BEST AVAILABLE COPY

PCT/NL 2004 / 0 0 0 2 2 9

21. 06. 04

Europäisches Patentamt

European Patent Office

Office européen des brevets

REC'D 08 JUL 2004

WIPO PCT

Bescheinigung

Certificate

Attestation

Die angehefteten Unterlagen stimmen mit der ursprünglich eingereichten Fassung der auf dem nächsten Blatt bezeichneten europäischen Patentanmeldung überein.

The attached documents are exact copies of the European patent application described on the following page, as originally filed.

Les documents fixés à cette attestation sont conformes à la version initialement déposée de la demande de brevet européen spécifiée à la page suivante.

Patentanmeldung Nr. Patent application No. Demande de brevet n°

03076110.0

PRIORITY DOCUMENT

SUBMITTED OR TRANSMITTED IN COMPLIANCE WITH RULE 17.1(a) OR (b)

Der Präsident des Europäischen Patentamts;
Im Auftrag

For the President of the European Patent Office

Le Président de l'Office européen des brevets
p.o.

R C van Dijk

Anmeldung Nr:
Application no.: 03076110.0
Demande no:

Anmeldetag:
Date of filing: 14.04.03
Date de dépôt:

Anmelder/Applicant(s)/Demandeur(s):

Vironovative B.V.
Burgemeester Oudlaan 50
3062 PA Rotterdam
PAYS-BAS

Bezeichnung der Erfindung/Title of the invention/Titre de l'invention:
(Falls die Bezeichnung der Erfindung nicht angegeben ist, siehe Beschreibung.
If no title is shown please refer to the description.
Si aucun titre n'est indiqué se referer à la description.)

Sars

In Anspruch genommene Prioriät(en) / Priority(ies) claimed /Priorité(s) revendiquée(s)
Staat/Tag/Aktenzeichen/State/Date/File no./Pays/Date/Numéro de dépôt:

EP/08.04.03/

Internationale Patentklassifikation/International Patent Classification/
Classification internationale des brevets:

C12N7/00

Am Anmeldetag benannte Vertragstaaten/Contracting states designated at date of filing/Etats contractants désignées lors du dépôt:

AT BE BG CH CY CZ DE DK EE ES FI FR GB GR HU IE IT LU MC NL
PT RO SE SI SK TR LI

P64721EP00

EPO - DG 1

14. 04. 2003

(65)

Title: SARS

The invention relates to the field of virology.

Recently, a new virus has caused a global health risk because of its pathogenic effects in man combined with a relatively easy droplet transmission. The virus first was seen in the Chinese province Guangdong, was spread to Hong Kong in February 2003, and within two months it has been able to spread to several countries all over the world where it has caused 78 deaths out of 2300 people infected (New Scientist Online News 13:25 02 April 2003). The virus has been named SARS (Severe Acute Respiratory Syndrome) virus and causes a respiratory illness (atypical pneumonia) in man. This illness usually begins with a fever, sometimes associated with chills or other symptoms, including headache, rash, diarrhea, a general feeling of discomfort (malaise) and body aches. Some people also experience mild respiratory syndromes at the outset.

After 2 to 7 days, SARS patients may develop a dry, nonproductive cough that might be accompanied or progress to the point where insuffiecient oxygen is getting to the blood, visible as shortness of breath. In 10% to 20% of the cases, patients will require mechanical ventilation, and eventually the disease can lead to the death of the patient. Hospital personnel, children, elderly and people having an underlying condition such as diabetes or heart disease, or a weakened immune system, form the highest risk group. Co-infection with other pathogens seems to occur frequently, especially with opportunistic pathogenic microorganisms such as human metapneumovirus (hMPV), Chlamydia, etcetera.

The incubation time for the virus is typically 2-7 days and the disease is transmitted by people sich with SARS coughing or sneezing droplets in the air.

As for yet it is not known if there is a cure for the disease. Several antiviral therapies have been applied, but with various results.

Also, for being able to prevent spread of the disease, it is of great importance to be able to recognise the disease in an early stage. Only then sufficient measures

can be taken to isolate patients and initiate quarantaine precautions. At this moment there is not yet a diagnostic tool in place.

Thus, there is great need in developing diagnostic tools and therapies for this disease.

The invention provides the nucleotide sequence of an isolated essentially mammalian positive-sense single stranded RNA virus belonging to the Coronaviruses, which is the causative factor for SARS. From a phylogenetic analysis of the sequences of the virus (Fig. 1) it appears that the virus is an intermediate between the group formed by TGEV (transmissable gastroenetritis virus), PEDV (porcine epidemic diarrhea virus) and 229E (human coronavirus 229E) at one side, the group formed by BoCo (bovine coronavirus) and MHV (murine hepatitis virus) at an other side, and the AIBV (avian infectious bronchitis virus) on yet another side . In general, bovine coronavirus seems to be the closest relative (at least for the viral replicase protein).

Although phylogenetic analyses provide a convenient method of identifying a virus as a SARS virus several other possibly more straightforward albeit somewhat more coarse methods for identifying said virus or viral proteins or nucleic acids from said virus are herein also provided. As a rule of thumb a SARS virus can be identified by the percentages of homology of the virus, proteins or nucleic acids to be identified in comparison with viral proteins or nucleic acids identified herein by sequence. It is generally known that virus species, especially RNA virus species, often constitute a quasi species wherein a cluster of said viruses displays heterogeneity among its members. Thus it is expected that each isolate may have a somewhat different percentage relationship with the sequences of the isolate as provided herein.

When one wishes to compare a virus isolate with the sequences as listed in figure 2, the invention provides an isolated essentially mammalian positive-sense single stranded RNA virus (SARS) belonging to the Coronaviruses and identifiable as phylogenetically corresponding thereto by determining a nucleic acid sequence of said virus and determining that said nucleic acid sequence has a percentage nucleic acid identity to the sequences as listed higher than the percentages identified herein for the nucleic acids as identified herein below in comparison with BoCo, AIPV and PEDV. Likewise, an isolated essentially mammalian positive-sense single stranded RNA virus (SARS) belonging to the Coronaviruses and identifiable as

phylogenetically corresponding thereto by determining an amino acid sequence of said virus and determining that said amino acid sequence has a percentage amino acid homology to the sequences as listed which is essentially higher than the percentages provided herein in comparison with BoCo, AIPV and PEDV.

With the provision of the sequence information of this SARS virus, the invention provides diagnostic means and methods, prophylactic means and methods and therapeutic means and methods to be employed in the diagnosis, prevention and/or treatment of disease, in particular of respiratory disease (atypical pneumonia), in particular of mammals, more in particular in humans. In virology, it is most advisory that diagnosis, prophylaxis and/or treatment of a specific viral infection is performed with reagents that are most specific for said specific virus causing said infection. In this case this means that it is preferred that said diagnosis, prophylaxis and/or treatment of a SARS virus infection is performed with reagents that are most specific for SARS virus. This by no means however excludes the possibility that less specific, but sufficiently cross-reactive reagents are used instead, for example because they are more easily available and sufficiently address the task at hand.
The invention for example provides a method for virologically diagnosing a SARS infection of an animal, in particular of a mammal, more in particular of a human being, comprising determining in a sample of said animal the presence of a viral isolate or component thereof by reacting said sample with a SARS specific nucleic acid or antibody according to the invention, and a method for serologically diagnosing a SARS infection of a mammal comprising determining in a sample of said mammal the presence of an antibody specifically directed against a SARS virus or component thereof by reacting said sample with a SARS virus-specific proteinaceous molecule or fragment thereof or an antigen according to the invention.
The invention also provides a diagnostic kit for diagnosing a SARS infection comprising a SARS virus, a SARS virus-specific nucleic acid, proteinaceous molecule or fragment thereof, antigen and/or an antibody according to the invention, and preferably a means for detecting said SARS virus, SARS virus-specific nucleic acid, proteinaceous molecule or fragment thereof, antigen and/or an antibody, said means for example comprising an excitable group such as a fluorophore or enzymatic detection system used in the art (examples of suitable diagnostic kit format comprise IF, ELISA, neutralization assay, RT-PCR assay). To determine whether an as yet

unidentified virus component or synthetic analogue thereof such as nucleic acid, proteinaceous molecule or fragment thereof can be identified as SARS-virus-specific, it suffices to analyse the nucleic acid or amino acid sequence of said component, for example for a stretch of said nucleic acid or amino acid, preferably of at least 10, more preferably at least 25, more preferably at least 40 nucleotides or amino acids (respectively), by sequence homology comparison with the provided SARS viral sequences and with known non-SARS viral sequences (BoCo is preferably used) using for example phylogenetic analyses as provided herein. Depending on the degree of relationship with said SARS or non-SARS viral sequences, the component or synthetic analogue can be identified.

The invention thus provides the nucleotide sequence of a novel etiological agent, an isolated essentially mammalian positive-sense single stranded RNA virus (herein also called SARS virus) belonging to the Coronaviridae family, and SARS virus-specific components or synthetic analogues thereof. Coronaviruses were first isolated from chickens in 1937, while the first human coronavirus was propagated *in vitro* by Tyrell and Bonoe in 1965. There are now about 13 species in this family, which infect cattle, pigs, rodents, cats, dogs, birds and man. Coronavirus particles are irregularly shaped, about 60-220 nm in diameter, with an outer envelope bearing distinctive, 'club-shaped' peplomers (about 20 nm long and 10 nm wide at the distal end). This 'crown-like' appearance give the family its name. The envelope carries two glycoproteins: S, the spike glycoprotein which is involved in cell fusion and is a major antigen, and M, the membrane glycoprotein, which is involved in budding and envelope formation. The genome is associated with a basic phosphoprotein, designated N. The genome of coronaviruses, a single stranded positive-sense RNA strand, is typically 27-31 Kb long and contains a 5' methylated cap and a 3' poly-A tail, by which it can directly function as an mRNA in the infected cell. Initially the 5' ORF 1 (about 20 Kb) is translated to produce a viral polymerase, which then produces a full length negative sense strand. This is used as a template to produce mRNA as a 'nested set' of transcripts, all with identical 5' non-translated leader sequence of 72 nucleotides and coincident 3' polyadenylated ends. Each mRNA thus produced is monocistronic, the genes at the 5' end being translated from the longest mRNA and so on. These unusual cytoplasmic structures are produced not by splicing, but by the polymerase during transcription. Between each of the genes there is a repeated

intergenic sequence – AACUAAAC – which interacts with the transcriptase plus cellular factors to splice the leader sequence onto the start of each ORF. In some coronaviruses there are about 8 ORFs, coding for the proteins mentioned above, but also for a heamagglutenin esterase (HE), and several other non-structural proteins. Newly isolated viruses are phylogenetically corresponding to and thus taxonomically corresponding to SARS virus when comprising a gene order and/or amino acid sequence and/or nucleotide sequence sufficiently similar to our prototypic SARS virus. The highest amino acid sequence homology, between SARS virus and any of the known other viruses of the same family to date (BoCo or Mouse Hapatitis Virus) is for parts of the polymerase protein 18-61% (the % homology, and the virus to which the homology is depend on the region of the polymerase that is examined), as can be deduced when comparing the sequences given in figure 2 with sequences of other viruses, in particular of BoCo and Mouse Hapatitis Virus. Individual proteins or whole virus isolates with, respectively, higher homology than these mentioned maximum values are considered phylogenetically corresponding and thus taxonomically corresponding to SARS virus, and generally will be encoded by a nucleic acid sequence structurally corresponding with a sequence as shown in figure 2. Herewith the invention provides a virus phylogenetically corresponding to the isolated virus of which the sequences are depicted in figure 2.

It should be noted that, similar to other viruses, a certain degree of variation can be expected to be found between SARS-viruses isolated from different sources.

Also, the viral sequence of the SARS virus or an an isolated SARS virus gene as provided herein for example shows less than 95%, preferably less than 90%, more preferably less than 80%, more preferably less than 70% and most preferably less than 65% nucleotide sequence homology or less than 95%, preferably less than 90%, more preferably less than 80%, more preferably less than 70% and most preferably less than 65% amino acid sequence homology with the respective nucleotide or amino acid sequence of the bovine coronavirus or the murine hepatitis virus as for example can be found in Genbank (for example in accession number NC_002306 (BoCo) or NC_002645 (MHV)).

Sequence divergence of SARS strains around the world may be somewhat higher, in analogy with other coronaviruses.

The term "nucleotide sequence homology" as used herein denotes the presence of homology between two (poly)nucleotides. Polynucleotides have "homologous"

sequences if the sequence of nucleotides in the two sequences is the same when aligned for maximum correspondence. Sequence comparison between two or more polynucleotides is generally performed by comparing portions of the two sequences over a comparison window to identify and compare local regions of sequence similarity. The comparison window is generally from about 20 to 200 contiguous nucleotides. The "percentage of sequence homology" for polynucleotides, such as 50, 60, 70, 80, 90, 95, 98, 99 or 100 percent sequence homology may be determined by comparing two optimally aligned sequences over a comparison window, wherein the portion of the polynucleotide sequence in the comparison window may include additions or deletions (i.e. gaps) as compared to the reference sequence (which does not comprise additions or deletions) for optimal alignment of the two sequences. The percentage is calculated by: (a) determining the number of positions at which the identical nucleic acid base occurs in both sequences to yield the number of matched positions; (b) dividing the number of matched positions by the total number of positions in the window of comparison; and (c) multiplying the result by 100 to yield the percentage of sequence homology. Optimal alignment of sequences for comparison may be conducted by computerized implementations of known algorithms, or by inspection. Readily available sequence comparison and multiple sequence alignment algorithms are, respectively, the Basic Local Alignment Search Tool (BLAST) (Altschul, S.F. et al. 1990. J. Mol. Biol. 215:403; Altschul, S.F. et al. 1997. Nucleic Acid Res. 25:3389-3402) and ClustalW programs both available on the internet. Other suitable programs include GAP, BESTFIT and FASTA in the Wisconsin Genetics Software Package (Genetics Computer Group (GCG), Madison, WI, USA).

As used herein, "substantially complementary" means that two nucleic acid sequences have at least about 65%, preferably about 70%, more preferably about 80%, even more preferably 90%, and most preferably about 98%, sequence complementarity to each other. This means that the primers and probes must exhibit sufficient complementarity to their template and target nucleic acid, respectively, to hybridise under stringent conditions. Therefore, the primer sequences as disclosed in this specification need not reflect the exact sequence of the binding region on the template and degenerate primers can be used. A substantially complementary primer sequence is one that has sufficient sequence complementarity to the amplification template to result in primer binding and second-strand synthesis.

The term "hybrid" refers to a double-stranded nucleic acid molecule, or duplex, formed by hydrogen bonding between complementary nucleotides. The terms "hybridise" or "anneal" refer to the process by which single strands of nucleic acid sequences form double-helical segments through hydrogen bonding between complementary nucleotides.

The term "oligonucleotide" refers to a short sequence of nucleotide monomers (usually 6 to 100 nucleotides) joined by phosphorous linkages (e.g., phosphodiester, alkyl and aryl-phosphate, phosphorothioate), or non-phosphorous linkages (e.g., peptide, sulfamate and others). An oligonucleotide may contain modified nucleotides having modified bases (e.g., 5-methyl cytosine) and modified sugar groups (e.g., 2'-O-methyl ribosyl, 2'-O-methoxyethyl ribosyl, 2'-fluoro ribosyl, 2'-amino ribosyl, and the like). Oligonucleotides may be naturally-occurring or synthetic molecules of double- and single-stranded DNA and double- and single-stranded RNA with circular, branched or linear shapes and optionally including domains capable of forming stable secondary structures (e.g., stem-and-loop and loop-stem-loop structures).

The term "primer" as used herein refers to an oligonucleotide which is capable of annealing to the amplification target allowing a DNA polymerase to attach thereby serving as a point of initiation of DNA synthesis when placed under conditions in which synthesis of primer extension product which is complementary to a nucleic acid strand is induced, i.e., in the presence of nucleotides and an agent for polymerization such as DNA polymerase and at a suitable temperature and pH. The (amplification) primer is preferably single stranded for maximum efficiency in amplification. Preferably, the primer is an oligodeoxy ribonucleotide. The primer must be sufficiently long to prime the synthesis of extension products in the presence of the agent for polymerization. The exact lengths of the primers will depend on many factors, including temperature and source of primer. A "pair of bi-directional primers" as used herein refers to one forward and one reverse primer as commonly used in the art of DNA amplification such as in PCR amplification.

The term "probe" refers to a single-stranded oligonucleotide sequence that will recognize and form a hydrogen-bonded duplex with a complementary sequence in a target nucleic acid sequence analyte or its cDNA derivative.

The terms "stringency" or "stringent hybridization conditions" refer to hybridization conditions that affect the stability of hybrids, e.g., temperature, salt concentration, pH, formamide concentration and the like. These conditions are empirically optimised

to maximize specific binding and minimize non-specific binding of primer or probe to its target nucleic acid sequence. The terms as used include reference to conditions under which a probe or primer will hybridise to its target sequence, to a detectably greater degree than other sequences (e.g. at least 2-fold over background). Stringent conditions are sequence dependent and will be different in different circumstances. Longer sequences hybridise specifically at higher temperatures. Generally, stringent conditions are selected to be about 5°C lower than the thermal melting point (Tm) for the specific sequence at a defined ionic strength and pH. The Tm is the temperature (under defined ionic strength and pH) at which 50% of a complementary target sequence hybridises to a perfectly matched probe or primer. Typically, stringent conditions will be those in which the salt concentration is less than about 1.0 M Na+ ion, typically about 0.01 to 1.0 M Na+ ion concentration (or other salts) at pH 7.0 to 8.3 and the temperature is at least about 30°C for short probes or primers (e.g. 10 to 50 nucleotides) and at least about 60°C for long probes or primers (e.g. greater than 50 nucleotides). Stringent conditions may also be achieved with the addition of destabilizing agents such as formamide. Exemplary low stringent conditions or "conditions of reduced stringency" include hybridization with a buffer solution of 30% formamide, 1 M NaCl, 1% SDS at 37°C and a wash in 2x SSC at 40°C. Exemplary high stringency conditions include hybridization in 50% formamide, 1 M NaCl, 1% SDS at 37°C, and a wash in 0.1x SSC at 60°C. Hybridization procedures are well known in the art and are described in e.g. Ausubel et al, Current Protocols in Molecular Biology, John Wiley & Sons Inc., 1994.

The term "antibody" includes reference to antigen binding forms of antibodies (e. g., Fab, F (ab) 2). The term "antibody" frequently refers to a polypeptide substantially encoded by an immunoglobulin gene or immunoglobulin genes, or fragments thereof which specifically bind and recognize an analyte (antigen). However, while various antibody fragments can be defined in terms of the digestion of an intact antibody, one of skill will appreciate that such fragments may be synthesized de novo either chemically or by utilizing recombinant DNA methodology. Thus, the term antibody, as used herein, also includes antibody fragments such as single chain Fv, chimeric antibodies (i. e., comprising constant and variable regions from different species), humanized antibodies (i. e., comprising a complementarity determining region (CDR) from a non-human source) and heteroconjugate antibodies (e. g., bispecific antibodies).

In short, the invention provides an isolated essentially mammalian positive-sense single stranded RNA virus (SARS) belonging to the Coronaviruses and identifiable as phylogenetically corresponding thereto by determining a nucleic acid sequence of a suitable fragment of the genome of said virus and testing it in phylogenetic tree analyses wherein maximum likelihood trees are generated using 100 bootstraps and 3 jumbles and finding it to be more closely phylogenetically corresponding to a virus isolate having the sequences as depicted in figure 2 than it is corresponding to a virus isolate of BoCo (bovine coronavirus, e.g. acc. no. NC_002306 in Genbank), MHV (murine hepatitis virus, e.g. acc. no. NC_002645), AIBV (avian infectious bronchitis virus, e.g. acc. no. NC_001451), PEDV (porcine epidemic diarrhea virus), TGEV (transmissible gastroenteritis virus, e.g. acc. no. NC_003436) or 229E (human coronavirus 229E, e.g. acc. no. NC_003045).

Suitable nucleic acid genome fragments each useful for such phylogenetic tree analyses are for example any of the RAP-PCR fragments EMC-1 to –14 and RDG-1 as disclosed in figure 2, leading to the phylogenetic tree analysis as disclosed herein in figure 1.

A suitable open reading frame (ORF) comprises the ORF encoding the viral polymerase (ORF 1a). When an overall amino acid identity of at least 60%, preferably of at least 70%, more preferably of at least 80%, more preferably of at least 90%, most preferably of at least 95% of the analysed polymerase with the polymerase having a sequence comprising the amino acid fragments EMC-1, EMC-2, EMC-3, EMC-4,EMC-5, EMC-13 and/or EMC-14 of figure 2 is found, the analysed virus isolate comprises a SARS virus isolate according to the invention.

Another suitable open reading frame (ORF) useful in phylogenetic analyses comprises the ORF encoding the N protein. When an overall amino acid identity of at least 60%, more preferably of at least70%, more preferably of at least 80%, more preferably of at least 90%, most preferably of at least 95% of the analysed N-protein with the N-protein encoded by a sequence comprising the sequence EMC-8 of figure 2 is found, the analysed virus isolate comprises a SARS isolate according to the invention.

Another suitable open reading frame (ORF) useful in phylogenetic analyses comprises the ORF encoding the spike protein S. When an overall amino acid identity of at least 60%, more preferably of at least 70%, more preferably of at least 80%, more

preferably of at least 90%, most preferably of at least 95%of the analysed S-protein encoded by a sequence comprising the sequence of translation 2 of EMC7 and translation 1 of the RDG 1 sequence of the S-protein as depicted in figure 2 is found, the analysed virus isolate comprises a SARS virus isolate according to the invention. The S ORF of the SARS virus seems to be located adjacent to the ORF 1ab (coding for the viral polymerase), which would discriminate SARS viruses from the bovine coronavirus and the murine hepatitis virus, which have a so-called 2a gene and an HE-gene between the S protein and the viral polymerase.

The invention provides among others an isolated or recombinant nucleic acid or virus-specific functional fragment thereof obtainable from a virus according to the invention. The isolated or recombinant nucleic acids comprises the sequences as given in figure 2 or sequences of homologues which are able to hybridise with those under stringent conditions. In particular, the invention provides primers and/or probes suitable for identifying a SARS virus nucleic acid.

Furthermore, the invention provides a vector comprising a nucleic acid according to the invention. To begin with, vectors such as plasmid vectors containing (parts of) the genome of SARS virus, virus vectors containing (parts of) the genome of SARS (for example, but not limited thereto, vaccinia virus, retroviruses, baculovirus), or SARS virus containing (parts of) the genome of other viruse or other pathogens are provided.

Also, the invention provides a host cell comprising a nucleic acid or a vector according to the invention. Plasmid or viral vectors containing the polymerase components of SARS virus are generated in prokaryotic cells for the expression of the components in relevant cell types (bacteria, insect cells, eukaryotic cells). Plasmid or viral vectors containing full-length or partial copies of the SARS virus genome will be generated in prokaryotic cells for the expression of viral nucleic acids *in-vitro* or *in-vivo*. The latter vectors may contain other viral sequences for the generation of chimeric viruses or chimeric virus proteins, may lack parts of the viral genome for the generation of replication defective virus, and may contain mutations, deletions or insertions for the generation of attenuated viruses.

Infectious copies of SARS virus (being wild type, attenuated, replication-defective or chimeric) can be produced upon co-expression of the polymerase components according to the state-of-the-art technologies described above.

In addition, eukaryotic cells, transiently or stably expressing one or more full-length or partial SARS virus proteins can be used. Such cells can be made by transfection (proteins or nucleic acid vectors), infection (viral vectors) or transduction (viral vectors) and may be useful for complementation of mentioned wild type, attenuated, replication-defective or chimeric viruses.

A chimeric virus may be of particular use for the generation of recombinant vaccines protecting against two or more viruses. For example, it can be envisaged that a SARS virus vector expressing one or more proteins of a human metapneumovirus or a human metapneumovirus vector expressing one or more proteins of SARS virus will protect individuals vaccinated with such vector against both virus infections. Such a specific chimeric virus is particularly useful in the invention because it is suspected that co-infection of, for instance, human metapneumovirus frequently occurs in SARS virus infetced patients. Attenuated and replication-defective viruses may be of use for vaccination purposes with live vaccines as has been suggested for other viruses.

In a preferred embodiment, the invention provides a proteinaceous molecule or coronavirus-specific viral protein or functional fragment thereof encoded by a nucleic acid according to the invention. Useful proteinaceous molecules are for example derived from any of the genes or genomic fragments derivable from a virus according to the invention. Such molecules, or antigenic fragments thereof, as provided herein, are for example useful in diagnostic methods or kits and in pharmaceutical compositions such as sub-unit vaccines and inhibitory peptides. Particularly useful are the viral polymerase protein, the spike protein, the nucleocapsid or antigenic fragments thereof for inclusion as antigen or subunit immunogen, but inactivated whole virus can also be used. Particulary useful are also those proteinaceous substances that are encoded by recombinant nucleic acid fragments that are identified for phylogenetic analyses, of course preferred are those that are within the preferred bounds and metes of ORFs useful in phylogenetic analyses, in particular for eliciting SARS virus specific antibodies, whether in vivo (e.g. for protective puposes or for providing diagnostic antibodies) or in vitro (e.g. by phage display technology or another technique useful for generating synthetic antibodies).

Also provided herein are antibodies, be it natural polyclonal or monoclonal, or synthetic (e.g. (phage) library-derived binding molecules) antibodies that specifically react with an antigen comprising a proteinaceous molecule or SARS virus-specific

functional fragment thereof according to the invention. Such antibodies are useful in a method for identifying a viral isolate as a SARS virus comprising reacting said viral isolate or a component thereof with an antibody as provided herein. This can for example be achieved by using purified or non-purified SARS virus or parts thereof (proteins, peptides) using ELISA, RIA, FACS or similar formats of antigen detection assays (Current Protocols in Immunology). Alternatively, infected cells or cell cultures may be used to identify viral antigens using classical immunofluorescence or immunohistochemical techniques. Specifically useful in this respect are antibodies raised against SARS virus proteins which are encoded by a nucleotide sequence comprising one or more of the fragments disclosed in figure 2.

Other methods for identifying a viral isolate as a SARS virus comprise reacting said viral isolate or a component thereof with a virus specific nucleic acid according to the invention.

In this way the invention provides a viral isolate identifiable with a method according to the invention as a mammalian virus taxonomically corresponding to a positive-sense single stranded RNA virus identifiable as likely belonging to the SARS virus genus within the family of Coronaviruses.

The method is useful in a method for virologically diagnosing a SARS virus infection of a mammal, said method for example comprising determining in a sample of said mammal the presence of a viral isolate or component thereof by reacting said sample with a nucleic acid or an antibody according to the invention.

Methods of the invention can in principle be performed by using any nucleic acid amplification method, such as the Polymerase Chain Reaction (PCR; Mullis 1987, U.S. Pat. No. 4,683,195, 4,683,202, en 4,800,159) or by using amplification reactions such as Ligase Chain Reaction (LCR; Barany 1991, Proc. Natl. Acad. Sci. USA 88:189-193; EP Appl. No., 320,308), Self-Sustained Sequence Replication (3SR; Guatelli et al., 1990, Proc. Natl. Acad. Sci. USA 87:1874-1878), Strand Displacement Amplification (SDA; U.S. Pat. Nos. 5,270,184, en 5,455,166), Transcriptional Amplification System (TAS; Kwoh et al., Proc. Natl. Acad. Sci. USA 86:1173-1177), Q-Beta Replicase (Lizardi et al., 1988, Bio/Technology 6:1197), Rolling Circle Amplification (RCA; U.S. Pat. No. 5,871,921), Nucleic Acid Sequence Based Amplification (NASBA), Cleavase Fragment Length Polymorphism (U.S. Pat. No. 5,719,028), Isothermal and Chimeric Primer-initiated Amplification of Nucleic Acid

(ICAN), Ramification-extension Amplification Method (RAM; U.S. Pat. Nos. 5,719,028 and 5,942,391) or other suitable methods for amplification of nucleic acids.

In order to amplify a nucleic acid with a small number of mismatches to one or more of the amplification primers, an amplification reaction may be performed under conditions of reduced stringency (e.g. a PCR amplification using an annealing temperature of 38°C, or the presence of 3.5 mM MgCl2). The person skilled in the art will be able to select conditions of suitable stringency.

The primers herein are selected to be "substantially" complementary (i.e. at least 65%, more preferably at least 80% perfectly complementary) to their target regions present on the different strands of each specific sequence to be amplified. It is possible to use primer sequences containing e.g. inositol residues or ambiguous bases or even primers that contain one or more mismatches when compared to the target sequence. In general, sequences that exhibit at least 65%, more preferably at least 80% homology with the target DNA or RNA oligonucleotide sequences, are considered suitable for use in a method of the present invention. Sequence mismatches are also not critical when using low stringency hybridization conditions.

The detection of the amplification products can in principle be accomplished by any suitable method known in the art. The detection fragments may be directly stained or labelled with radioactive labels, antibodies, luminescent dyes, fluorescent dyes, or enzyme reagents. Direct DNA stains include for example intercalating dyes such as acridine orange, ethidium bromide, ethidium monoazide or Hoechst dyes. Alternatively, the DNA or RNA fragments may be detected by incorporation of labelled dNTP bases into the synthesized fragments. Detection labels which may be associated with nucleotide bases include e.g. fluorescein, cyanine dye or BrdUrd.

When using a probe-based detection system, a suitable detection procedure for use in the present invention may for example comprise an enzyme immunoassay (EIA) format (Jacobs et al., 1997, J. Clin. Microbiol. 35, 791-795). For performing a detection by manner of the EIA procedure, either the forward or the reverse primer used in the amplification reaction may comprise a capturing group, such as a biotin group for immobilization of target DNA PCR amplicons on e.g. a streptavidin coated microtiter plate wells for subsequent EIA detection of target DNA -amplicons (see below). The skilled person will understand that other groups for immobilization of target DNA PCR amplicons in an EIA format may be employed.

Probes useful for the detection of the target DNA as disclosed herein preferably bind only to at least a part of the DNA sequence region as amplified by the DNA amplification procedure. Those of skill in the art can prepare suitable probes for detection based on the nucleotide sequence of the target DNA without undue experimentation as set out herein. Also the complementary nucleotide sequences, whether DNA or RNA or chemically synthesized analogs, of the target DNA may suitably be used as type-specific detection probes in a method of the invention, provided that such a complementary strand is amplified in the amplification reaction employed.

Suitable detection procedures for use herein may for example comprise immobilization of the amplicons and probing the DNA sequences thereof by e.g. southern blotting. Other formats may comprise an EIA format as described above. To facilitate the detection of binding, the specific amplicon detection probes may comprise a label moiety such as a fluorophore, a chromophore, an enzyme or a radio-label, so as to facilitate monitoring of binding of the probes to the reaction product of the amplification reaction. Such labels are well-known to those skilled in the art and include, for example, fluorescein isothiocyanate (FITC), β-galactosidase, horseradish peroxidase, streptavidin, biotin, digoxigenin, 35S or 125I. Other examples will be apparent to those skilled in the art.

Detection may also be performed by a so called reverse line blot (RLB) assay, such as for instance described by Van den Brule et al. (2002, J. Clin. Microbiol. 40, 779-787). For this purpose RLB probes are preferably synthesized with a 5' amino group for subsequent immobilization on e.g. carboxyl-coated nylon membranes. The advantage of an RLB format is the ease of the system and its speed, thus allowing for high throughput sample processing.

The use of nucleic acid probes for the detection of RNA or DNA fragments is well known in the art. Mostly these procedure comprise the hybridization of the target nucleic acid with the probe followed by post-hybridization washings. Specificity is typically the function of post-hybridization washes, the critical factors being the ionic strength and temperature of the final wash solution. For nucleic acid hybrids, the Tm can be approximated from the equation of Meinkoth and Wahl, Anal. Biochem., 138: 267-284 (1984): $Tm = 81.5\ °C + 16.6 (\log M) + 0.41 (\%\ GC) - 0.61 (\%\ form) - 500/L$; where M is the molarity of monovalent cations, % GC is the percentage of guanosine and cytosine nucleotides in the nucleic acid, % form is the percentage of formamide in the

hybridization solution, and L is the length of the hybrid in base pairs. The Tm is the temperature (under defined ionic strength and pH) at which 50% of a complementary target sequence hybridizes to a perfectly matched probe. Tm is reduced by about 1 °C for each 1 % of mismatching; thus, the hybridization and/or wash conditions can be adjusted to hybridize to sequences of the desired identity. For example, if sequences with > 90% identity are sought, the Tm can be decreased 10°C. Generally, stringent conditions are selected to be about 5 °C lower than the thermal melting point (Tm) for the specific sequence and its complement at a defined ionic strength and pH. However, severely stringent conditions can utilize a hybridization and/or wash at 1,2,3, or 4 °C lower than the thermal melting point (Tm); moderately stringent conditions can utilize a hybridization and/or wash at 6,7,8,9, or 10 °C lower than the thermal melting point (Tm); low stringency conditions can utilize a hybridization and/or wash at 11,12,13,14,15, or 20 °C lower than the thermal melting point (Tm). Using the equation, hybridization and wash compositions, and desired Tm, those of ordinary skill will understand that variations in the stringency of hybridization and/or wash solutions are inherently described. If the desired degree of mismatching results in a Tm of less than 45 °C (aqueous solution) or 32 °C (formamide solution) it is preferred to increase the SSC concentration so that a higher temperature can be used. An extensive guide to the hybridization of nucleic acids is found in Tijssen, Laboratory Techniques in Biochemistm and Molecular Biology—Hybridization with Nucleic Acid Probes, Part I, Chapter 2" Overview of principles of hybridization and the strategy of nucleic acid probe assays", Elsevier. New York (1993); and Current Protocols in Molecular Biology, Chapter 2, Ausubel, et al., Eds., Greene Publishing and Wiley-Interscience, New York (1995).

In another aspect, the invention provides oligonucleotide probes for the generic detection of target RNA or DNA. The detection probes herein are selected to be "substantially" complementary to one of the strands of the double stranded nucleic acids generated by an amplification reaction of the invention. Preferably the probes are substantially complementary to the immobilizable, e.g. biotin labelled, antisense strands of the amplicons generated from the target RNA or DNA.

It is allowable for detection probes of the present invention to contain one or more mismatches to their target sequence. In general, sequences that exhibit at least 65%, more preferably at least 80% homology with the target oligonucleotide sequences are considered suitable for use in a method of the present invention.

Antibodies, both monoclonal and polyclonal, can also be used for detection purpose in the present invention, for example, in immunoassays in which they can be utilized in liquid phase or bound to a solid phase carrier. In addition, the monoclonal antibodies in these immunoassays can be detectably labeled in various ways. A variety of immunoassay formats may be used to select antibodies specifically reactive with a particular protein (or other analyte). For example, solid-phase ELISA immunoassays are routinely used to select monoclonal antibodies specifically immunoreactive with a protein. See Harlow and Lane, Antibodies, A Laboratory Manual, Cold Spring Harbor Publications, New York (1988), for a description of immunoassay formats and conditions that can be used to determine selective binding. Examples of types of immunoassays that can utilize antibodies of the invention are competitive and non-competitive immunoassays in either a direct or indirect format. Examples of such immunoassays are the radioimmunoassay (RIA) and the sandwich (immunometric) assay. Detection of the antigens using the antibodies of the invention can be done utilizing immunoassays that are run in either the forward, reverse, or simultaneous modes, including immunohistochemical assays on physiological samples. Those of skill in the art will know, or can readily discern, other immunoassay formats without undue experimentation.

Antibodies can be bound to many different carriers and used to detect the presence of the target molecules. Examples of well-known carriers include glass, polystyrene, polypropylene, polyethylene, dextran, nylon, amylases, natural and modified celluloses, polyacrylamides, agaroses and magnetite. The nature of the carrier can be either soluble or insoluble for purposes of the invention. Those skilled in the art will know of other suitable carriers for binding monoclonal antibodies, or will be able to ascertain such using routine experimentation.

The invention also provides a method for serologically diagnosing a SARS virus infection of a mammal comprising determining in a sample of said mammal the presence of an antibody specifically directed against a SARS virus or component thereof by reacting said sample with a proteinaceous molecule or fragment thereof or an antigen according to the invention

Methods and means provided herein are particularly useful in a diagnostic kit for diagnosing a SARS virus infection, be it by virological or serological diagnosis. Such kits or assays may for example comprise a virus, a nucleic acid, a proteinaceous

molecule or fragment thereof, an antigen and/or an antibody according to the invention.

Use of a virus, a nucleic acid, a proteinaceous molecule or fragment thereof, an antigen and/or an antibody according to the invention is also provided for the production of a pharmaceutical composition, for example for the treatment or prevention of SARS virus infections and/or for the treatment or prevention of atypical pneumonia, in particular in humans. Preferably a peptide comprising part of the amino acid sequence of the spike protein as depicted in translation 2 with the sequence EMC7 and translation 1 of the RDG seq of figure 2, is used for the preparation of a therapeutic or prophylactic peptide. Also preferably, a protein comprising the amino acid sequence of the spike protein as depicted in translation 2 with the sequence EMC7 translation 1 of the RDG seqof figure 2, is used for the preparation of a sub-unit vaccine. Furthermore, the nucleocapsid of Cornoviruses, as depicted in the translation of EMC8, in figure 2, is known to be particularly useful for eliciting cell-mediated immunity against Coronaviruses and can be used for the preparation of a sub-unit vaccine.

Attenuation of the virus can be achieved by established methods developed for this purpose, including but not limited to the use of related viruses of other species, serial passages through laboratory animals or/and tissue/cell cultures, serial passages through cell cultures at temparutes below 37C (cold-adaption), site directed mutagenesis of molecular clones and exchange of genes or gene fragments between related viruses.

A pharmaceutical composition comprising a virus, a nucleic acid, a proteinaceous molecule or fragment thereof, an antigen and/or an antibody according to the invention can for example be used in a method for the treatment or prevention of a SARS virus infection and/or a respiratory illness comprising providing an individual with a pharmaceutical composition according to the invention. This is most useful when said individual comprises a human. Antibodies against SARS virus proteins, especially against the spike protein of SARS virus, preferably against the amino acid sequence as depicted in translation 2 of EMC7 and translation 1 of the RDG seq in figure 2, are also useful for prophylactic or therapeutic purposes, as passive vaccines. It is known from other coronaviruses that the spike protein is a very strong antigen and that antibodies against spike protein can be used in prophylactic and therapeutic vaccination.

The invention also provides method to obtain an antiviral agent useful in the treatment of atypical pneumonia comprising establishing a cell culture or experimental animal comprising a virus according to the invention, treating said culture or animal with an candidate antiviral agent, and determining the effect of said agent on said virus or its infection of said culture or animal. An example of such an antiviral agent comprises a SARS virus-neutralising antibody, or functional component thereof, as provided herein, but antiviral agents of other nature are obtained as well. The invention also provides use of an antiviral agent according to the invention for the preparation of a pharmaceutical composition, in particular for the preparation of a pharmaceutical composition for the treatment of atypical pneumonia, especifically when caused by a SARS virus infection, and provides a pharmaceutical composition comprising an antiviral agent according to the invention, useful in a method for the treatment or prevention of a SARS virus infection or atypical pneumonia, said method comprising providing an individual with such a pharmaceutical composition.

The invention also comprises an animal model usable for testing of prophylactic and/or therapeutic methods and/or preparations. It has appeared that apes can be infected with the SARS virus, thereby showing clinical symptoms, and more importantly, similar tissue morphology as found in humans suffering from atypical pneumonia caused by the SARS virus. Subjecting apes to a prophylactic or therapeutic treatment either before or during infection with the virus will have a good and useful predictionary value for application of such a prophylaxis or therapy in human subjects.

The invention is further explained in the Examples without limiting it thereto.

Figure legends

Fig. 1: Phylogenetic relationship for the nucleotide sequences of isolate HK39849 with its closest relatives genetically. Phylogenetic trees were generated by maximum likelihood analyses using 100 bootstraps and 3 jumbles. The scale representing the number of nucleotide changes is shown for each tree.

Fig. 2: Nucleotide sequences from 13 clones of parts of the SARS virus. Also included are the putative polypeptide sequences of polypeptides and alignments of the putative polypeptides with that of another member of the Coronoviridae family, where possible.

Fig. 3: Schematic map of the SARS virus genome, indicating the position of the nucleotide sequences of figure 2 relative to the genome and a putative indication of the open reading frames of the genome based on analogy with other coronaviruses. The gene structure for the region between the Spike and Nucleocapsid is uncertain. EMC1-EMC14 and RDG 1: sequences as provided in figure 2. CDC and BIN1-2: sequences were provided through personal communication from the CDC (Dr. W. Bellini, Centers for Disease Control & Prevention, National Centers for Infectious Diseases, 1600 Clifton Road, Atlanta GA 30333, USA) and BNI (Dr. C . Drosten and Prof. Dr. H. Schmitz, Bernard Nocht Institute, Bernard-Nocht Str. 74, D-20359 Hamburg, Germany), respectively.

Fig. 4: Amino acid comparison of the N-terminus of the S-protein of the SARS virus and closely related coronaviruses. HCV OC43 = human coronavirus isolate OC43; MHV A59 = murine hepatitis virus isolate A59, BCV = bovine corona virus.

Fig. 5: Negative contrast EM photograph of SARS virus obtained from concentrated supernatant of infected cell cultures.

Fig. 6: Infection with SARS-coronavirus causes pulmonary and renal lesions in cynomolgus macaques. Formalin-fixed, paraffin-embedded tissues were stained with haematoxylin and eosin and examined by light microscopy. There is diffuse alveolar damage of the lung (a), and the alveolar lumina (b) are flooded with highly

proteinaceous exudate admixed with inflammatory cells and cellular debris. In the lumen of a bronchiole (c) and in the surrounding lung parenchyma are several multinucleated syncytial cells (arrowheads). The renal collecting tubules (d) contain similar multinucleated syncytial cells. Original magnifications: a x 12.5; b x 50; c x 100; d x 250.

Examples

Virus isolation and characterisation

Isolate HK39849 was isolated from a hospitalised SARS patient by throat swab and inoculated into a culture of Vero-E6 cells. A sample of the supernatant from these infected cells was provided by Dr. M. Peiris (Queeen Mary Hospital Faculty of Medicine, Hong Kong University, Honk Kong) was used to inoculate VERO-118 cells and cell culture supernatant from these cells was aliquoted and frozen after one passage

We isolated RNA from the virus-containing cell culture supernatant and subjected it to RNA arbitrarily primed PCR (RAP-PCR) essentially as described by Welsh & McClelland (NAR 18:7213; PNAS USA 90:10710, 1993). Virus in the culture supernatants was purified on continuous 20-60% sucrose gradients. The gradient fractions were inspected for virus-like particles by EM, and RNA was isolated from the fraction containing , in which the most nucleocapsids were observed. Equivalent amounts of RNA isolated from virus fractions were used for RAP-PCR, after which samples were run side by side on a 3% NuSieve agarose gel. Differentially displayed bands ranging in size from 200-1500 base pairs specific for the unidentified virus were subsequently purified from the gel, cloned in plasmid pCR2.1 (Invitrogen) and sequenced with vector-specific primers. When we used these sequences to search for homologies against sequences in the Genbank database using the BLAST software (www.ncbi.nlm.nih.gov/BLAST/) which yielded resemblance to virus sequences of the coronaviruses displayed in the phylogenetic tree of figure 1.

Eight of these fragments (EMC 1-6, 13 and 14) were located in the ORF coding for the viral polymerase (ORF 1ab), one (EMC-7) spanned the 3' end of ORF1ab and reached into the 5' end of spike protein region; EMC-10 overlapped the 3' end of EMC-7 and therefore also codes part of the S protein region and EMC 9 encodes a region downstream of EMC-10; by use of primers to sequences within EMC10 and EMC9 (see below), the region between these two sequences was amplified by PCR and sequenced. The full contiguous region has been incoporated into EMC7 in firgure 2;a further sequence (RDG1 in figure 2) encodes the 3' end of the Spike protein. A further sequence (EMC8) spanned part of the Nucleocapsid coding sequence. The remaining three sequences (EMC9, 11 and 12) encode regions of as yet unknown function.

Phylogeny

BLAST searches using nucleotide sequences obtained from the unidentified virus isolate revealed homologies primarily with members of the Coronaviridae. As an indication for the relation between the newly identified virus isolate and other coronaviruses a phylogenetic tree was constructed based on the sequence information obtained (figure 1).

Materials and Methods

Specimen collection

Virus was collected from SARS patients using throat swabs and from experimentally infected monkeys (throat and nasal swabs, serum, plasma and faeces)

Virus isolation and culture

Throat swabs were dipped into a culture of Vero-E6 cells and incubated for 1-4 days. Cell culture supernatant was clarified by centrifugation and filtered through a 0.45micrometre filter, before beings stored frozen. The virus was subsequently propagated in Vero-118 cells.

Antigen detection by indirect IFA

Samples from experimentally infected monkeys was cultured on Vero-118 cells in 24 well plates containing glass slides. These glass slides were washed with PBS and fixed in aceton for 1 minute at room temperature. After washing with PBS the slides were incubated for 30 minutes at 37 °C with SARS-antibody containing serum from SARS patients. After washing off the human serum in PBS, the slides were incubated at 37°C for 30 minutes with FITC labeled anti-human antibodies. After three washes in PBS and one in tap water, the slides were included in a glycerol/PBS solution (Citifluor, UKC, Canterbury, UK) and covered. The slides were analysed using an Axioscop fluorescence microscope (Carl Zeiss B.V., Weesp, the Netherlands).

Detection of antibodies in humans by indirect IFA

Virus was cultured on Vero-118 cells in 24 well plates containing glass slides. These glass slides were washed with PBS and fixed in aceton for 1 minute at room temperature. After washing with PBS the slides were incubated for 30 minutes at 37

°C with SARS-antibody containing serum from SARS patients. After washing off the human serum in PBS, the slides were incubated at 37°C for 30 minutes with FITC labeled anti-human antibodies. After three washes in PBS and one in tap water, the slides were included in a glycerol/PBS solution (Citifluor, UKC, Canterbury, UK) and covered. The slides were analysed using an Axioscop fluorescence microscope (Carl Zeiss B.V., Weesp, the Netherlands

Detection of antibodies in humans byELISA

Patient samples.

4 samples of patients with SARS disease , 8 samples of patients from routine serological virology; samples from an experimentally infected monkey (preserum, 9 and 12 days after infection).

The Conjugate.

Whole virus was used as the conjugate.. Tissue culture supernatant from infected Vero cells were pelleted through 20% sucrose onto a 60% sucrose cushion. The virus was then pelleted through 20% sucrose and resuspended in PBS/1% NP40. After dialysis using PBS, the virus was The conjugated to horseradish peroxidase by standard techniques was tested in 3 concentrations (diluted in dilution buffer 9000-03, 1:100, 1:400 and 1:1600), both on polyvalent anti-IgM code MCB0201 (cross-reactive with monkey) and monoclonal anti-IgM, code 9000-62 (non-crossreactive with monkey).

Sera were diluted 1:200 in serum diluent (code 9000-03), monkey 775 was diluted 1: 100, 1:200 and 1:400.

Serum incubation one hour at 37°C, conjugate incubation one hour at 37°C, and TMB (ready to use): 30 minutes at room temperature. The reaction was stopped with sulphuric acid (0.5M).

Virus characterisation

For EM analyses, virus was concentrated from infected cell culture supernatants in a micro-centrifuge at 4 °C at 17000 x g, after which the pellet was resuspended in PBS and inspected by negative contrast EM

RNA isolation

RNA was isolated from the supernatant of infected cell cultures or sucrose gradient fractions using a High Pure RNA Isolation kit according to instructions from the manufacturer (Roche Diagnostics, Almere, The Netherlands).

RT-PCR

A one-step RT-PCR was performed in 50 µl reactions containing 50 mM Tris.HCl pH 8.5, 50 mM NaCl, 4 mM MgCl2, 2 mM dithiotreitol, 200 µM each dNTP, 10 units recombinant RNAsin (Promega, Leiden, the Netherlands), 10 units AMV RT (Promega, Leiden, The Netherlands), 5 units Amplitaq Gold DNA polymerase (PE Biosystems, Nieuwerkerk aan de Ijssel, The Netherlands) and 5 µl RNA. Cycling conditions were 45 min. at 42 °C and 7 min. at 95 °C once, 1 min at 95 °C, 2 min. at 42 °C and 3 min. at 72 °C repeated 40 times and 10 min. at 72 °C once.

Primers used for diagnostic PCR:

SARS fwd2: ggtggaacatcatccggtgat

SARS rev2: agcctgtgttgtagattgcgg

These primers amplify a 149bp fragment of the polymerase gene (orf 1ab)

RF 999: TTTAAACACTTACGAGAGTTTGTG

RF997: GGACACAACCCATGAAATCATCTGG

These primers amplify a region of 728bp in the spike glycoprotein gene (S)

RF998: AGACATATCTAATGTGCCTTTCTCC RF1002:

AAGCTCGTCACCTAAGTCATAAGAC (from EMC11 sequence)

The combination of RF998/RF1002 primers enabled us to sequence the 3' end of EMC7 – RF998 is a specific primer withing EMC7 whereas EMC1002 acted as a random primer.

RT-PCR, gel purification and direct sequencing were performed as described above.

RAP-PCR

RAP-PCR was performed essentially as described by Welsh & McClelland (Nuc. Acid Res. 18:7213, 1990; Proc. Natl. Acad. Sci. USA 90:10710 1993) . The oligonucleotide sequences are described in addenda 2. For the RT reaction, 2 µl RNA was used in a 10 µl reaction containing 10 ng/µl oligonucleotide, 10 mM dithiotreitol, 500 µm each

dNTP, 25 mM Tris-HCl pH 8.3, 75 mM KCl and 3 mM MgCl2. The reaction mixture was incubated for 5 min. at 70 °C and 5 min. at 37 °C, after which 200 units Superscript RT enzyme (LifeTechnologies) were added. The incubation at 37 °C was continued for 55 min. and the reaction terminated by a 5 min. incubation at 72 °C. The RT mixture was diluted to give a 50 µl PCR reaction containing 8 ng/µl oligonucleotide, 300 µm each dNTP, 15 mM Tris-HCL pH 8.3, 65 mM KCl, 3.0 mM $MgCL_2$ and 5 units Taq DNA polymerase (PE Biosystems). Cycling conditions were 5 min. at 94 °C, 5 min. at 40 °C and 1 min. at 72 °C once, followed by 1 min. at 94 °C, 2 min. at 56 °C and 1 min. at 72 °C repeated 40 times and 5 min. at 72°C once. After RAP-PCR, 15 µl the RT-PCR products were run side by side on a 3% NuSieve agarose gel (FMC BioProducts, Heerhugowaard, The Netherlands). Differentially displayed fragments were purified from the gel with Qiaquick Gel Extraction kit (Qiagen, Leusden, The Netherlands) and cloned in pCR2.1 vector (Invitrogen, Groningen, The Netherlands) according to instructions from the manufacterer.

Sequence analysis

RAP-PCR products cloned in vector pCR2.1 (Invitrogen) were sequenced with M13-specific oligonucleotides. DNA fragments obtained by RT-PCR were purified from agarose gels using Qiaquick Gel Extraction kit (Qiagen, Leusden, The Netherlands), and sequenced directly with the same oligonucleotides used for PCR. Sequence analyses were performed using a Dyenamic ET terminator sequencing kit (Amersham Pharmacia Biotech, Roosendaal, The Netherlands) and an ABI 373 automatic DNA sequencer (PE Biosystem). All techniques were performed according to the instructions of the manufacturer.

RT-PCR for diagnosing SARS virus.

For the amplification of the SARS virus' genetic material, we used primers:

SARS fwd2: ggtggaacatcatccggtgat

SARS rev2: agcctgtgttgtagattgcgg

These primers amplify a 149bp fragment of the polymerase gene (orf 1ab)

RF 999: TTTAAACACTTACGAGAGTTTGTG

RF997: GGACACAACCCATGAAATCATCTGG

These primers amplify a region of 728bp in the spike glycoprotein gene (S)

These primers amplify a 149bp fragment of the polymerase gene (orf 1ab) RT-PCR, gel purification and direct sequencing were performed as described above.

Phylogenetic analyses

For all phylogenetic trees, DNA sequences were alligned using the ClustalW software package and maximum likelihood trees were generated using the DNA-ML software package of the Phylip 3.5 program using 100 bootstraps and 3 jumbles[15]. Previously published sequences for TGEV, PEDV, 229E, AIBV, BoCo and MHV that were used for the generation of phylogenetic trees are available from Genbank

Examples of methods to identify SARS virus

Specimen collection

In order to find virus isolates nasopharyngeal aspirates, throat and nasal swabs, broncheo alveolar lavages, serum and plasma samples, and stools preferably from mammals such as humans, carnivores (dogs, cats, mustellits, seals etc.), horses, ruminants (cattle, sheep, goats etc.), pigs, rabbits, birds (poultry, ostriches, etc) should be examined. From birds cloaca swabs and droppings can be examined as well. Sera should be collected for immunological assays, such as ELISA, molecular-based assays, such as RT-PCR and virus neutralisation assays.
Collected virus specimens were diluted with 5 ml Dulbecco MEM medium (BioWhittaker, Walkersville, MD) and thoroughly mixed on a vortex mixer for one minute. The suspension was thus centrifuged for ten minutes at 840 x g. The sediment was spread on a multispot slide (Nutacon, Leimuiden, The Netherlands) for immunofluorescence techniques, and the supernatant was used for virus isolation.

Virus isolation

For virus isolation Vero-118 cells or tMK cells (RIVM, Bilthoven, The Netherlands) were cultured in 24 well plates containing glass slides (Costar, Cambridge, UK), with the medium described below supplemented with 10% fetal bovine serum (BioWhittaker, Vervier, Belgium). Before inoculation the plates were washed with PBS and supplied with Eagle's MEM with Hanks' salt (ICN, Costa mesa, CA)

supplemented with 0.52/liter gram $NaHCO_3$, 0.025 M Hepes (Biowhittaker), 2 mM L-glutamine (Biowhittaker), 200 units/liter penicilline, 200 µg/liter streptomycine (Biowhittaker), 1gram/liter lactalbumine (Sigma-Aldrich, Zwijndrecht, The Netherlands), 2.0 gram/liter D-glucose (Merck, Amsterdam, The Netherlands), 10 gram/liter peptone (Oxoid, Haarlem, The Netherlands) and 0.02% trypsine (Life Technologies, Bethesda, MD). The plates were inoculated with supernatant of the patient samples, 0,2 ml per well in triplicate, followed by centrifuging at 840x *g* for one hour. After inoculation the plates were incubated at 37 °C for a maximum of 1-3 days and cultures were checked daily for CPE. Extensive CPE was generally observed within 24hours. and included detachment of cells from the monolayer..

Virus culture of SARS

Sub-confluent monolayers of tMK cells or Vero clone 118 cells in media as described above were inoculated with supernatants of samples that displayed CPE or with samples taken from patient or artificially infected monkeys..

Virus characterisation

For EM analyses, virus was concentrated from infected cell culture supernatants in a micro-centrifuge at 4 °C at 17000 x g, after which the pellet was resuspended in PBS and inspected by negative contrast EM.

Antigen detection by indirect IFA

Virus was cultured on Vero-118 cells in 24 well slides containing glass slides. These glass slides were washed with PBS and fixed in aceton for 1 minute at room temperature.

After washing with PBS the slides were incubated for 30 minutes at 37 °C with SARS patient serum. We used patient serum, but antibodies can be raised in various animals, such as ferrets, goats and rabbits (for polyclonal antibodies) and mice and hamsters (for monoclonal antibodies), and the working dilution of the antibody can vary for each immunisation. After three washes with PBS and one wash with tap water, the slides were incubated at 37°C for 30 minutes with FITC labeled goat-anti-human antibodies. After three washes in PBS and one in tap water, the slides were included in a glycerol/PBS solution (Citifluor, UKC, Canterbury, UK) and covered.

The slides were analysed using an Axioscop fluorescence microscope (Carl Zeiss B.V., Weesp, the Netherlands).

Detection of antibodies in humans by indirect IFA

For the detection of virus specific antibodies, SARS virus-infected Vero cells were fixed with acetone on coverslips (as described above), washed with PBS and incubated 30 minutes at 37°C with serum samples at a 1 to 16 dilution. After two washes with PBS and one with tap water, the slides were incubated 30 minutes at 37°C with FITC-labelled secondary antibodies to human antibodies (Dako). Slides were processed as described above.

Antibodies can be labelled directly with a fluorescent dye, which will result in a direct immuno fluorescence assay. FITC can be replaced with any fluorescent dye. This technique can be applied to antibodies in other animals such as mammals, ruminants, birds or other species, assuming the secondary antibody to the appropriate species is used.

Detection of antibodies in humans by ELISA

Patient samples.

4 samples of patients with SARS; 8 samples of patients from routine serological virology; samples from an experimentally infected monkey (preserum and 9 days after infection).

The Conjugate.

The conjugate was tested at a number of concentrations, both on polyvalent anti-IgM (cross-reactive with monkey) and monoclonal anti-IgM, (non-crossreactive with monkey).

Sera were diluted 1:200 in serum diluent and the monkey serum was diluted 1: 100, 1:200 and 1:400.

Serum incubation one hour at 37°C, conjugate incubation one hour at 37°C, and TMB (ready to use): 30 minutes at room temperature. The reaction was stopped with sulphuric acid (0.5M).

Results were interpreted by eye. Three of the four SARS-IgM positive sera (as detected by IF on infected cells) had a higher score than negative control sera.One serum had a score which was also reached by some of the negative controls. The 9 day old monkey sera did not react, but the 12 day old did. Thus, this study shows that with direct conjugation of nucleocapsids the developemnt of an IgM capture method is feasable.

Furthermore, this type of assay can be performed in a number of formats by those trained in the art. The assay can be extended to the detection of IgA and IgG antibodies from humans and animals and can make use of different capture antigens, such as, but not limited to, purified recombinant N protein.

Animal immunisation

Cynomologous macaque specific antisera for the newly discovered virus were generated by experimental intratrachael installation of cultured virus of Cynomologous macaques. One to two weeks later the animals were bled. The sera were tested for reactivity to SARS virus by indirect IFA as described above; uninfected control cells were used to ensure the specificity of the serum. Other animal species are also suitable for the generation of specific antibody preparations and other antigen preparations may be used.

RNA isolation

RNA was isolated from the supernatant of infected cell cultures or sucrose gradient fractions using a High Pure RNA Isolation kit according to instructions from the manufacturer (Roche Diagnostics, Almere, The Netherlands). RNA can also be isolated following other procedures known in the field (*Current Protocols in Molecular Biology*).

RT-PCR

A one-step RT-PCR was performed in 50 μl reactions containing 50 mM Tris.HCl pH 8.5, 50 mM NaCl, 4 mM $MgCl_2$, 2 mM dithiotreitol, 200 μM each dNTP, 10 units recombinant RNAsin (Promega, Leiden, the Netherlands), 10 units AMV RT (Promega, Leiden, The Netherlands), 5 units Amplitaq Gold DNA polymerase (PE Biosystems, Nieuwerkerk aan de Ijssel, The Netherlands) and 5 μl RNA. Cycling

conditions were 45 min. at 42 °C and 7 min. at 95 °C once, 1 min at 95 °C, 2 min. at 42 °C and 3 min. at 72 °C repeated 40 times and 10 min. at 72 °C once.

Primers used for diagnostic PCR:

For the amplification of the SARS virus' genetic material, we used primers:

SARS fwd2: ggtggaacatcatccggtgat

SARS rev2: agcctgtgttgtagattgcgg

These primers amplify a 149bp fragment of the polymerase gene (orf 1ab)

RT-PCR, gel purification and direct sequencing were performed as described above.

Sequence analysis

Sequence analyses were performed using a Dyenamic ET terminator sequencing kit (Amersham Pharmacia Biotech, Roosendaal, The Netherlands) and an ABI 373 automatic DNA sequencer (PE Biosystem). All techniques were performed according to the instructions of the manufacturer. PCR fragments were sequenced directly with the same oligonucleotides used for PCR, or the fragments were purified from the gel with Qiaquick Gel Extraction kit (Qiagen, Leusden, The Netherlands) and cloned in pCR2.1 vector (Invitrogen, Groningen, The Netherlands) according to instructions from the manufacturer and subsequently sequenced with M13-specific oligonucleotides.

Detection of antibodies in humans, mammals, ruminants or other animals by ELISA

A recombinant protein derived from the SARS virus is preferred as the antigen. However, purified nucleocapsids may also be used. Antigens suitable for antibody detection include any SARS protein that combines with any SARS-specific antibody of a patient exposed to or infected with SARS virus. Preferred antigens of the invention include those that predominantly engender the immune response in patients exposed to SARS, which therefore, typically are recognised most readily by antibodies of a patient. Particularly preferred antigens include the N, and S proteins of SARS.

Antigens used for immunological techniques can be native antigens or can be modified versions thereof. Well known techniques of molecular biology can be used to

alter the amino acid sequence of a SARS antigen to produce modified versions of the antigen that may be used in immunologic techniques.

Methods for cloning genes, for manipulating the genes to and from expression vectors, and for expressing the protein encoded by the gene in a heterologous host are well-known, and these techniques can be used to provide the expression vectors, host cells, and the for expressing cloned genes encoding antigens in a host to produce recombinant antigens for use in diagnostic assays. See for instance: *Molecular cloning, A laboratory manual* and *Current Protocols in Molecular Biology*.

A variety of expression systems may be used to produce SARS antigens. For instance, a variety of expression vectors suitable to produce proteins in *E.Coli*, *B.subtilis*, yeast, insect cells and mammalian cells have been described, any of which might be used to produce a SARS antigen suitable to detect anti- SARS antibodies in exposed patients.

The baculovirus expression system has the advantage of providing necessary processing of proteins, and is therefor preferred. The system utilizes the polyhedrin promoter to direct expression of SARS antigens. (Matsuura et al. 1987, J.Gen.Virol. 68: 1233-1250).

Antigens produced by recombinant baculo-viruses can be used in a variety of immunological assays to detect anti- SARS antibodies in a patient. It is well established, that recombinant antigens can be used in place of natural virus in practically any immunological assay for detection of virus specific antibodies.

The assays include direct and indirect assays, sandwich assays, solid phase assays such as those using plates or beads among others, and liquid phase assays. Assays suitable include those that use primary and secondary antibodies, and those that use antibody binding reagents such as protein A. Moreover, a variety of detection methods can be used in the invention, including colorimetric, fluorescent, phosphorescent, chemiluminescent, luminescent and radioactive methods.

Animal model example

Four Cynomologous Macaques were infected with SARS virus by intratrachaeal installation using Vero-118 cell derived virus.

The monkeys had the following clinical symptoms

- Lethargy
- One of four monkeys had severe pneumonia
- Mild to severe rash in the inguinal region and the axilar region
- Watery stools

After 10-16 days the monkeys were euthanized. Tissues were examined and the following was found

- Alveolae were filled with serum and their architecture were disrupted, consistent with bronchointestitial pneumonia (see fig 5 and b)
- Multi-cell syncytia in lungs (fig 5c)
- Multi-cell syncytia in kidneys (fig 5d)
- Widening of the small intestine

Virus was detected using RT-PCR on tissue samples and by culturing samples followed by electron microscopy from

- Lungs
- Nasal swabs
- Throat swabs
- Faeces
- Kidneys

The EM results demonstrate that the virus that was recovered from the Cynomologous Macaques had the identical morphology to the virus which was used to infect them.

This demonstrates that Cynomologous Macaques may be used as animal models to tests the efficacy of pharmaceutical preparations for therapeutic or prophylactic purposes

Claims

1. An isolated essentially mammalian positive-sense single stranded RNA virus (SARS) comprising one or more of the sequences of figure 2.

2. An isolated positive-sense single stranded RNA virus (SARS) belonging to the Coronaviruses and identifiable as phylogenetically corresponding thereto by determining a nucleic acid sequence of said virus and testing it in phylogenetic tree analyses wherein maximum likelihood trees are generated using 100 bootstraps and 3 jumbles and finding it to be more closely phylogenetically corresponding to a virus isolate having the sequences as depicted in figure 2 than it is corresponding to a virus isolate of BoCo (bovine coronavirus), MHV (murine hepatitis virus), AIBV (avian infectious bronchitis virus), PEDV (porcine epidemic diarrhea virus), TGEV (transmissible gastroenteritis virus) or 229E (human coronavirus 229E)..

3. A virus according to claim 1 or 2 wherein said nucleic acid sequence comprises an open reading frame (ORF) encoding a viral protein of said virus.

4. A virus according to claim 3 wherein said open reading frame is selected from the group of ORFs encoding the viral replicase, nuclear capsid protein and the spike protein.

5. A virus according to claim 1-4 isolatable from a human with atypical pneumonia.

6. An isolated or recombinant nucleic acid or SARS virus-specific functional fragment thereof obtainable from a virus according to anyone of claims 1 to 5.

7. A vector comprising a nucleic acid according to claim 6.

8. A host cell comprising a nucleic acid according to claim 6 or a vector according to claim 7.

9. An isolated or recombinant proteinaceous molecule or SARS virus-specific functional fragment thereof encoded by a nucleic acid according to claim 6.

10. An antigen comprising a proteinaccous molecule or SARS virus-specific functional fragment thereof according to claim 9.

11. An antibody specifically directed against an antigen according to claim 10.

12. A method for identifying a viral isolate as a SARS virus comprising reacting said viral isolate or a component thereof with an antibody according to claim 11.

13. A method for identifying a viral isolate as a SARS virus comprising reacting said viral isolate or a component thereof with a nucleic acid according to claim 6.

14. A method for virologically diagnosing a SARS infection of a mammal comprising determining in a sample of said mammal the presence of a viral isolate or component thereof by reacting said sample with a nucleic acid according to claim 6 or an antibody according to claim 11.

15. A method for serologically diagnosing a SARS infection of a mammal comprising determining in a sample of said mammal the presence of an antibody specifically directed against a SARS virus or component thereof by reacting said sample with a proteinaceous molecule or fragment thereof according to claim 9 or an antigen according to claim 10.

16. A diagnostic kit for diagnosing a SARS infection comprising a virus according to anyone of claims 1 to 5, a nucleic acid according to claim 6, a proteinaceous molecule or fragment thereof according to claim 9, an antigen according to claim 10 and/or an antibody according to claim 11.

17. Use of a virus according to any one claims 1 to 5, a nucleic acid according to claim 6, a vector according to claim 7, a host cell according to claim 8, a proteinaceous molecule or fragment thereof according to claim 9, an antigen according to claim 10,

or an antibody according to claim 11 for the production of a pharmaceutical composition.

18. Use according to claim 17 for the production of a pharmaceutical composition for the treatment or prevention of a SARS virus infection.

19. Use according to claim 17 or 18 for the production of a pharmaceutical composition for the treatment or prevention of atypical pneumonia.

20. A pharmaceutical composition comprising a virus according to any one of claims 1 to 5, a nucleic acid according to claim 6, a vector according to claim 7, a host cell according to claim 8, a proteinaceous molecule or fragment thereof according to claim 9, an antigen according to claim 10, or an antibody according to claim 11.

21. A method for the treatment or prevention of a SARS virus infection comprising providing an individual with a pharmaceutical composition according to claim 20.

22. A method for the treatment or prevention of atypical pneumonia comprising providing an individual with a pharmaceutical composition according to claim 20.

23. A viral replicase encoded by an RNA sequence comprising the sequences EMC-1, EMC-2, EMC-3, EMC-4, EMC-5, EMC-6, EMC-7, EMC-13 and/or EMC-14, or homologues thereof as depicted in figure 2.

24. A viral spike protein comprising the amino acid depicted as translation 2 with sequence EMC-7 and translation 1 of RDG 1 as depicted in figure 2, or a homologue thereof.

25 A viral nuclear capsid protein encoded by an RNA sequence comprising the sequence EMC-8 as depicted in figure 2 or a homologue thereof.

26. A viral protein encoded by an RNA sequence comprising the sequence EMC-9, EMC-11 and/or EMC-12 as depicted in figure 2.

27. A nucleic acid sequence which comprises one or more of the sequences EMC-1 to EMC-13 as depicted in figure 13 or a nucleic acid sequence which can hybridise with any of these sequences under stringent conditions.

Abstract

The invention relates to the field of virology. The invention provides an isolated essentially mammalian positive-sense single stranded RNA virus (SARS) within the group of coronaviuses and components thereof.

Figure 1.

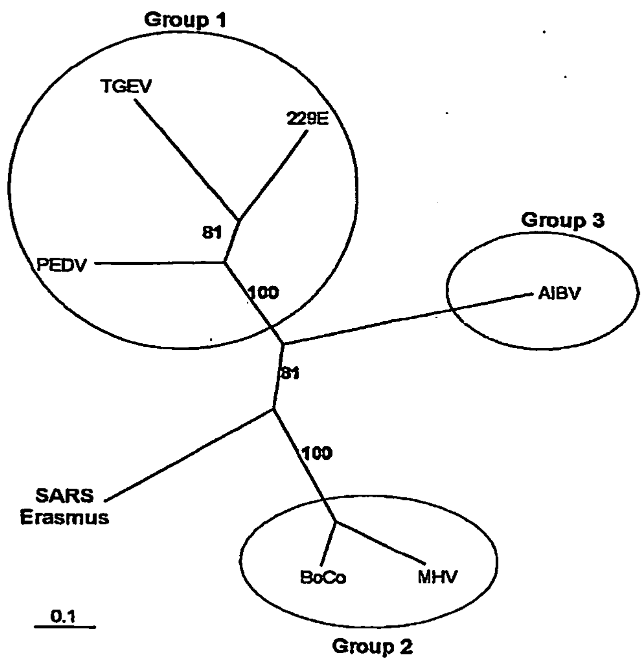
Group 1
TGEV
229E
81
PEDV
100
Group 3
AIBV
81
100
SARS
Erasmus
BoCo
MHV
0.1
Group 2

Figure 2 RNA sequences, implied polypeptides and alignment with one close relative

EMC-1

UUGUAACUGGUGGUCUUGUACAACAGACUUCUCAGUGGUUGUCUAAUCUUUUGGCACUACUGGUUGAAAAAC
UCAGGCCUAUCUUUGAAUGGAUUGAGGCGAAACUUAGUGCAGGAGUUGAAUUUCUCAAGGAUGCUUGGGAGAU
UCUCAAAUUUCUCAUUACAGGUGUUUUUGACAUCGUCAAGGGUCAAAUACAGGUUGCUUCAGAUAACAUCAAG
GAUUGUGUAAAAUGCUUCAUUGAUGUUGUUAACAAGGCACUCGAAAUGUGCAUUGAUCAAGUCACUAUCGCUG
GCGCAAAGUUGCGAUCACUCAACUUAGGUGAAGUCUUCAUCGCUCAAAGCAAGGGACUUUACCGUCAGUGUAU
ACGUGGCAAGGAGCAGCUGCAACUACUCAUGCCUCUUAAGGCACCAAAAGAAGUAACCUUUCU
UGAAGGUGAUUCACAUGACACAGUACUUACCUCUGAGGAGGUUGUUCUCAAGAACGGUGAA
CUCGAAGCACUCGAGACGCCCGUUGAUAGCUUCACAAAUGGAGCUAUCGUUGGCACACCAG
UCUGUGUAAAUGGCCUCAUGCUCUUAGAGAUUAAGGACAAAGAACAAUACUGCGCAUUGUC
UCCUGGUUUACUGGCUACaAACAAUGUCUUUCGCUUAAAAGGGGGUGCACCAAUUAAAGGU
GUAACCUUUGGAGAAGAUACUGUUUGGGAAGUUCAGGGUUACAAGAAUGUGAGAAUCACAU
UUGAGCUUGAUGAACGUGUUGACAAAGUGCUUAAUGAAAAGUGCUCUGUCUACACUGUUGA
AUCCGGUACCGAAGUUACUGAGUUUGCAUGUGUUGUAGCAGAGGCUGUUGUGAAGACUUUA
CAACCAGUUUCUGAUC

*Translation*Nucleotides 7 to 870: Frame 1; 288 aa

LVVLYNRLLSGCLIFWALLVEKLRPIFEWIEAKLSAGVEFLKDAWEILKFLITGVFDIVKGQIQVASDNIKDCVKCFIDVV
NKALEMCIDQVTIAGAKLRSLNLGEVFIAQSKGLYRQCIRGKEQLQLLMPLKAPKEVTFLEGDSHDTVLTSEEVVLKNGEL
EALETPVDSFTNGAIVGTPVCVNGLMLLEIKDKEQYCALSPGLLATNNVFRLKGGAPIKGVTFGEDTVWEVQGYKNVRITF
ELDERVDKVLNEKCSVYTVESGTEVTEFACVVAEAVVKTLQPVSD

Alignment

```
RNA-directed RNA polymerase (orf1a) murine hepatitis virus
Identities = 72/285 (25%), Positives = 118/285 (41%)

Query: 49    FWALLVEKLRPIFEWIEAKLSAGVEFLKDAWEILKFLITGVFDIVKGQIQVASDNIKDCV 228
             F AL V  +R I EW +  L+  +      W   + L+ G+F +  G I    + + + V
Sbjct: 638   FKALGVAVVRKITEWFD--LAVDIAASAAGWLCYQ-LVNGLFAVANGVITFVQE-VPELV 693

Query: 229   KCFIDVVNKALEMCIDQVTIA---GAKLRSLNLGEVFIAQSKGLYRQCIRGKEQLQLLMP 399
             K F+D      ++ ID ++++   G  +       V +A SK +Y   +  K     +MP
Sbjct: 694   KNFVDKFKAFFKVLIDSMSVSILSGLTVVKTASNRVCLAGSK-VYE--VVQKSLSAYVMP 750

Query: 400   LKAPKEVTFLEGDSHDTVLTSEEVVLKNGEL--EALETPVDSFTNGAIVGTPVCVNGLML 573
             +    E T L G+    V   + V +    L  +    P  SF    IV          L
Sbjct: 751   VGC-SEATCLVGEIEPAVFEDDVVDVVKAPLTYQGCCKPPTSFEKICIVDK--------L 801

Query: 574   LEIKDKEQYCAL-----SPGLLATNNVFRLKGGAPIKGVTFGEDT-VWEVQGYKNVRITF 735
                K  +Q+  +     + G+L     F   G    K V F +   V ++   + ++ITF
Sbjct: 802   YMAKCGDQFYPVVVDNDTVGVLDQCWRFPCAG----KKVEFNDKPKVRKIPSTRKIKITF 857

Query: 736   ELDERVDKVLNEKCSVYTVESGTEVTEFACVVAEAVVKTLQPVSD 870
              LD   D VL++ CS + V+    + E   VV +AV  TL P  +
Sbjct: 858   ALDATFDSVLSKACSEFEVDKDVTLDELLDVVLDAVESTLSPCKE 902
```

EMC-14

CAUCCAGCUUCUUAAGGCAGCAUAUGAAAAUUUCAAUUCACAGGACAUCUUACUUGCACCAUUGUUGUCAGCA
GGCAUAUUUGGUGCUAAACCACUUCAGUCUUUACAAGUGUGCGUGCAGACGGUUCGUACACAGGUUUAUAUUG
CAGUCAAUGACAAAGCUCUUUAUGAGCAGGUUGUCAUGGAUUAUCUUGAUAACCUGAAGCCUAGAGUGGAAGC
ACCUAAACAAGAGGAGCCACCAAACACAGAAGAUUCCAAAACUGAGGAGAAAUCUGUCGUACAGAAGCCUGUC
GAUGUGAAGCCAAAAAUUAAGGCCUGCAUUGAUGAGGUUACCACAACACUGGAAGAAACUAAGUUUCUUACCA
AUAAGUUACUCUUGUUUGCUGAUAUCAAUGGUAAGCUUUACCAUGAUUCUCAGAACAUGCUUAGAGGUGAAGA
UAUGUCUUUCCUUGAGAAGGAUGCACCUUACAUGGUAGGUGAUGUUAUCACUAGUGGUGAUAUCACUUGUGUU

Fig. 2, Contd.

GUAAUACCCUCCAAAAAGGCUGGUGGCACUACUGAGAUGCUCUCAAGAGCUUUGAAGAAAGUGCCAGUUGAUG
AGUAUAUAACCACGUACCCUGGACAAGGAUGUGCUGGUUAUACACUUGAGGAAGCUAAGACUGCUCUUAAGAA
AUGCAAAUCUGCAUUUUAUGUACUACCUUCAGAAGCACCUAAUGCUAAGGAAGAGAUUCUAGGAACUGUAUCC
UGGAAUUGAG

Translation

Nucleotides 5 to 739: Frame 2; 245 aa
IQLLKAAYENFNSQDILLAPLLSAGIFGAKPLQSLQVCVQTVRTQVYIAVNDKALYEQVVMDYLDNLKPRVEAPKQEEPPN
TEDSKTEEKSVVQKPVDVKPKIKACIDEVTTTLEETKFLTNKLLLFADINGKLYHDSQNMLRGEDMSFLEKDAPYMVGDVI
TSGDITCVVIPSKKAGGTTEMLSRALKKVPVDEYITTYPGQGCAGYTLEEAKTALKKCKSAFYVLPSEAPNAKEEILGTVS
WN

Alignment

replicase polyprotein 1ab Human coronavirus 229E

Identities = 48/202 (23%), Positives = 83/202 (41%), Gaps = 13/202 (6%)
Frame = +2

```
Query: 8    LLKAAYENFNSQDILLAPLLSAGIFGAKPLQSLQVCVQTVRT---QVYIAVNDKALYEQV 178
            L+KA     N Q   L P+LS GIFG K    SL+V +     T     +V++   + +     +
Sbjct: 1371 LIKAYNTINNEQGTPLTPILSCGIFGIKLETSLEVLLDVCNTKEVKVFVYTDTEVCKVKD 1430

Query: 179  VMDYLDNLKPRVEAPKQEEPPNTEDSKTEEKSVVQKPVDVKPKIKACIDEVTTTLEETKF 358
             +  L N++ +VE PK E   P                V  KP  V  K        +++      ++
Sbjct: 1431 FVSGLVNVQ-KVEQPKIEPKP------VSVIKVAPKPYRVDGKFSYFTEDLLCVADDKPI 1483

Query: 359  L--TNKLLLFADINGKLYHDSQNMLRG--EDMSFLEKDAP------YMVGDVITSGDITC 508
            +  T+ +L    D     L +      +L     +D    + K   P          + +G V+       +
Sbjct: 1484 VLFTDSMLTLDDRGLALDNALSGVLSAAIKDCVDINKAIPSGNLIKFDIGSVV----VYM 1539

Query: 509  VVIPSKKAGGTTEMLSRALKKV 574
             V+PS+K          + R  +K+
Sbjct: 1540 CVVPSEKDKHLDNNVQRCTRKL 1561
```

EMC-2

UCGAGAUUUcAUcUUGACGGUGCAGGUUCUUUCACUUGACAAACUAAAGAGUCUCUUAUCCCUGCGGGAGGUU
AAGACUAUAAAAGUGUUCACAACUGUGGACAACACUAAUCUCCACACACAGCUUGUGGAUAUGUCUAUGACAU
AUGGACAGCAGUUUGGUCCAACAUACUUGGAUGGUGCUGAUGUUACAAAAAUUAAACCUCAUGUAAAUCAUGA
GGGUAAGACUUUCUUUGUACUACCUAGUGAUGACACACUACGUAGUGAAGCUUUCGAGUACUACCAUACUCUU
GAUGAGAGUUUUCUUGGUAGGUACAUGUCUGCUUUAAACCACACAAAGAAAUGGAAA

Translation

Nucleotide 2 to 349: Frame 2; 116 aa

RDFILTVQVLSLDKLKSLLSLREVKTIKVFTTVDNTNLHTQLVDMSMTYGQQFGPTYLDGADVTKIKPHVNHEGKTFFVLP
SDDTLRSEAFEYYHTLDESFLGRYMSALNHTKKWK

Alignment

> Bovine Coronavirus RNA-Dependent RNA polymerase

Identities = 25/90 (27%), Positives = 44/90 (48%)
Frame = +2

```
Query: 80   IKVFTTVDNTNLHTQLVDMSMTYGQQFGPTYLDGADVTKIKPHVNHEGKTFFVLPSDDTL 259
            + +  TVD  N    + V +  ++G+  G  + DG +VTK K  +N++GK FF    +  +
Sbjct: 1565 VDILLTVDGVNFTNRFVPVGESFGKSLGNVFCDGVNVTKHKCDINYKGKVFFQFDNLSSE 1624

Query: 260  RSEAFEYYHTLDESFLGRYMSALNHTKKWK 349
             +A          D+  L  Y + L +  KW+
Sbjct: 1625 DLKAVRSSFNFDQKELLAYYNMLVNCSKWQ 1654
```

Fig.2, Côntd.

EMC13:

CUGAAGAAGUAGUGGaAAAUCCUACCAUACAGAAGGAAGUCAUAGAGUGUGACGUGAAAACUACCGAAGUUGU
AGGCAAUGUCAUACUUAAACCAUCAGAUGAAGGUGUUAAAGUAACACAAGAGUUAGGUCAUGAGGAUCUUAUG
GCUGCUUAUGUGGAAAACACAAGCAUUACCAUUAAGAAACCUAAUGAGCUUUCACUAGCCUUAGGUUUAAAAA
CAAUUGCCACUCAUGGUAUUGCUGCAAUUAAUAGUGUUCCUUGGAGUAAAAUUUUGGCUUAUGUCAAACCAUU
CUUAGGACAAGCAGCAAUUACAACAUCAAAUUGCGCUAAGAGAUUAGCACAACGUGUGUUUAACAAUUAUAUG
CCUUAUGUGUUUACAUUAUUGUUCCAAUUGUGUACUUUUACUAAAAGUACCAAUUCUAGAAUUAGAGCUUCAC
UACCUACAACUAUUGCUAAAAAUAGUGUUAAGAGUGUUGCUAAAUUAUGUUUGGAUGCCGGCAUUAAUUAUGU
GAAGUCACCCAAAUUUUCUAAAUUGUUCACAAUCGCUAUGUGGCUAUUGUUGUUAAGUAUUUGCUUAGGUUCU
CUAAUCUGUGUAACUGCUGCUUUUGGUGUACUCUUAUCUAAUUUUGGUGCUCCUUCUUAUUGUAAUGGCGUUA
GAGAAUUGUAUCUUAAUUCGUCUAACGUUACUACUAUGGAUUUCUGUGAAGGUUCUUUUCCUUGCAGCAUUUG
UUUAAGUGGAUUAGACUCCCUUGAUUCUUAUCCAGCUCUUGAAACCAUUCAGGUGACGAUUUCAUCGUACAAG
CUAGACUUGACAAUUUUAGGUCUGGCCGCUG

Translation

>~out: 3 to 833: Frame 3 277 aa
EEVVENPTIQKEVIECDVKTTEVVGNVILKPSDEGVKVTQELGHEDLMAAYVENTSITIKKPNELSLALGLKTIATHGIAA
INSVPWSKILAYVKPFLGQAAITTSNCAKRLAQRVFNNYMPYVFTLLFQLCTFTKSTNSRIRASLPTTIAKNSVKSVAKLC
LDAGINYVKSPKFSKLFTIAMWLLLLSICLGSLICVTAAFGVLLSNFGAPSYCNGVRELYLNSSNVTTMDFCEGSFPCSIC
LSGLDSLDSYPALETIQVTISSYKLDLTILGLAA

Alignment

bovine coronavirus RNA-dependent RNA Polymerase
Identities = 50/269 (18%),

```
Query: 57    KTTEVVGNVILKPSDEGVKVTQELGHEDLMAAYVENTSITIKKPNELSLALGLKTIATH- 233
             K  +V  +VI+       +K  + L   D+   ++        ++  N+LS+A+ + TI
Sbjct: 2046 KPFKVEDSVIVNDDTSEIKYVKSLSIVDVYDMWLTGCRYVVRTANDLSMAVNVPTIRKFI 2105

Query: 234   --GIAAINSVPWSKI-LAYVKPFLGQAAITTSNCAKRLAQRVFN--NYMPYVFTLLF--- 389
               G+   + S+P    + L  +KP            N  K +  ++    N++ ++F LLF
Sbjct: 2106 KFGMTLV-SIPIDLLNLREIKPVF--------NVVKAVRNKISACFNFIKWLFVLLFGWI 2156

Query: 390   -----QLCTFTKSTNSRIRASLPTTIAKNSVKSVAKLCLDAGINYVKSPKFSKLFTIAMW 554
                   +T    S++    L      KN+  +       +  G    + +          +W
Sbjct: 2157 KISADNKVIYTTEVASKLTCKLVALAFKNAFLTFKWSVVARGACIIAT-------IFLLW 2209

Query: 555   XXXXXXXXXXXXVTAAFGVLLSNFGAPSYCNGVRELYLNSSNVTTM------------- 695
                                G L      P++    + +    ++ ++ T+
Sbjct: 2210 FNFIYANVIFSDFYLPKIGFL------PTFVGKIAQWIKSTFSLVTICDLYSIQDVGFKN 2263

Query: 696   DFCEGSFPCSICLSGLDSLDSYPALETIQ 782
               +C GS  C  CL+G D LD+Y A++ +Q
Sbjct: 2264 QYCNGSIACQFCLAGFDMLDNYKAIDVVQ 2292
```

EMC-3

GUGGUAAGAUUGUUAGUACUUGUUUUAAACUUAUGCUUAAGGCCACAUUAUUGUGCGUUCU
UGCUGCAUUAGUUUGUUAUAUCGUUAUGCCAGUACAUACAUUGUCAAUCCAUGAUGGUUAC
ACAAAUGAAAUCAUUGGUUACAAAGCCAUUCAGGAUGGUGUCACUCGUGACAUCAUUUCUA
CUGAUGAUUGUUUUGCAAAUAAACAUGCUGGUUUUGACGCAUGGUUUAGCCAGCGUGGUGG
UUCAUACAAAAAUGACAAAAGCUGCCCUGUAGUAGCUGCUAUCAUUACAAGAGAGAUUGGU
UUCAUAGUGCCUGGCUUACCGGGUACUGUGCUGAGAGCAAUCAAUGGUGACUUCUUGCAUU
UCCUACCUCGUGUUUUUAGUGCUGUUGGCAACAUUUGCUACACACCUUCCAAACUCAUUGA
GUAUAGUGAUUUUGCUACCUCU

Translation

Nucleotide 3-449; 149 aa
GKIVSTCFKLMLKATLLCVLAALVCYIVMPVHTLSIHDGYTNEIIGYKAIQDGVTRDIISTDDCFANKHAGFD
AWFSQRGGSYKNDKSCPVVAAIITREIGFIVPGLPGTVLRAINGDFLHFLPRVFSAVGNICYTPSKLIEYSDF
ATS

Fig. 2, Contd.

Alignment

> Murine Hepatitis Virus RNA-Dependent RNA polymerase

Identities = 48/126 (38%),

```
Query: 78   YIVMPVHTLSIHDGYTNEIIGYKAIQDGVTRDIISTDDCFANKHAGFDAWFSQRGG--SY 251
            + +MP + +   D          +K I +GV RD+  TD CFANK   FD W+    G   Y
Sbjct: 2859 WALMPTYAVHKSDMQLPLYASFKVIDNGVLRDVSVTDACFANKFNQFDQWYESTFGLAYY 2918

Query: 252  KNDKSCPVVAAIITREIGFIVPGLPGTVLRAINGDFLHFLPRVFSAVGNICYTPSKLIEY 431
            +N K+CPVV A+I ++IG  +  +P TVLR       LHF+   F+      CYTP   I Y
Sbjct: 2919 RNSKACPVVVAVIDQDIGHTLFNVPTTVLR-YGFHVLHFITHAFATDSVQCYTPHMQIPY 2977

Query: 432  SDFATS 449
             +F  S
Sbjct: 2978 DNFYAS 2983
```

<u>EMC-4</u>

ACAGACAUCAAUCACUUCUGCUGUUCUGCAGAGUGGUUUUAGGAAAAUGGCAUUCCCGUCAGGCAAAGUUGAA
GGGUGCAUGGUACAAGUAACCUGUGGAACUACAACUCUUAAUGGAUUGUGGUUGGAUGACACAGUAUACUGUC
CAAGACAUGUCAUUUGCACAGCAGAAGACAUGCUUAAUCCUAACUAUGAAGAUCUGCUCAUUCGCAAAUCCAA
CCAUAGCUUUCUUGUUCAGGCUGGCAAUGUUCAACUUCGUGUUAUUGGCCAUUCUAUGCAAAAUUGUCUGCUU
AGGCUUAAAGUUGAUACUUCUAACCCUAAGACACCCAAGUAUAAAUUUGUCCGUAUCCAACCUGGUCAAACAU
UUUCAGUUCUAGCAUGCUACAAUGGUUCACCAUCUGGUGUUUAUCAGUGUGCCAUGAGACCUAAUCAUACCAU
UAAAGGUUCUUUCCUUAAUGGAUCAUGUGGUAGUGUUGGUUUUAACAUUGAUUAUGAUUGCGUGUCUUUCUGC
UAUAUGCAUCAUAUGGAGCUUCCAACAGGAGUACACGCUGGUACUGACUUAGAAGGUAAAUUCUAUGGUCCAU
UUGUUGACAGACAAACUGCACAGGCUGCAGGUACAGACACAACCAUAACAUUAAAUGUUUUGGCAUGGCUGUA
UGCUGCUGUUAUCAAUGGUGAUA

Translation

Nucleotides 2 to 679: Frame 2; 226 aa
QTSITSAVLQSGFRKMAFPSGKVEGCMVQVTCGTTTLNGLWLDDTVYCPRHVICTAEDMLNPNYEDLLIRKSNHSFLVQAG
NVQLRVIGHSMQNCLLRLKVDTSNPKTPKYKFVRIQPGQTFSVLACYNGSPSGVYQCAMRPNHTIKGSFLNGSCGSVGFNI
DYDCVSFCYMHHMELPTGVHAGTDLEGKFYGPFVDRQTAQAAGTDTTITLNVLAWLYAAVINGD

Alignment

RNA-directed RNA polymerase murine hepatitis virus

Identities = 122/222 (54%)

```
Query: 8    SITSAVLQSGFRKMAFPSGKVEGCMVQVTCGTTTLNGLWLDDTVYCPRHVICTAEDMLNP 187
            S+T++ LQSG  KM  P+ KVE C+V VT G  TLNGLWLDD VYCPRHVIC++ DM +P
Sbjct: 3326 SVTTSFLQSGIVKMVSPTSKVEPCIVSVTYGNMTLNGLWLDDKVYCPRHVICSSADMTDP 3385

Query: 188  NYEDLLIRKSNHSFLVQAGNVQLRVIGHSMQNCLLRLKVDTSNPKTPKYKFVRIQPGQTF 367
            +Y +LL R ++  F V +G + L V+ + MQ C L L V   NP TPKY F  ++PG+TF
Sbjct: 3386 DYPNLLCRVTSSDFCVMSGRMSLTVMSYQMQGCQLVLTVTLQNPNTPKYSFGVVKPGETF 3445

Query: 368  SVLACYNGSPSGVYQCAMRPNHTIKGSFLNGSCGSVGFNIDYDCVSFCYMHHMELPTGVH 547
            +VLA YNG P G +   +R +HTIKGSFL GSCGSVG+ +   D V F YMH +EL TG H
Sbjct: 3446 TVLAAYNGRPQGAFHVTLRSSHTIKGSFLCGSCGSVGYVLTGDSVRFVYMHQLELSTGCH 3505

Query: 548  AGTDLEGKFYGPFVDRQTAQAAGTDTTITLNVLAWLYAAVIN 673
             GTD  G FYGP+ D Q  Q    D T T+NV+AWLYAA+ N
Sbjct: 3506 TGTDFSGNFYGPYRDAQVVQLPVQDYTQTVNVVAWLYAAIFN 3547
```

<u>EMC-5</u>

Note that this sequence is not fully in frame.
AGUUGGAAAAGAUGGCAGAUCAGGCUAUGACCCAAAUGUACAAACAGGCAAGAUCUGAGGA
CAAGAGGGCAAAAGUAACUAGUGCUAUGCAAACAAUGCUCUUCACUAUGCUUAGGAAGCUU
GAUAAUGAUGCACUUAACAACAUUAUCAACAAUGCGCGUGAUGGUUGUGUUCCACUCAACA
UCAUACCAUUGACUACAGCAGCCAAACUCAUGGUUGUUGUCCCUGAUUAUGGUACCUACAA
GAACACUUGUGAUGGUAACACCUUUACAUAUGCAUCUGCACUCUGGGAAAUCCAGCAAGUU
GUUGAUGCGGAUAGCAAGAUUGUUCAACUUAGUGAAAUUAACAUGGACAAUUCACCAAAUU
UGGCUUGGCCCCUUAUUGUUACAGCUCUAAGAGCCAACUCAGCUGUUAAACUACAGAAUAA
UGAACUGAGUCCAGUAGCACUACGACAGAUGUCCUGUGCGGCUGGUACCACACAAACAGCU
UGUACUGAUGACAAUGCACUUGCCUACUAUAACAAUUCGAAGGGAGGUAGGUUUGUGCUGG

Fig. 2, Contd.

CAUUACUAUCAGACCACCAAGAUCUCAAAUGGGCUAGAUUCCCUAAGAGUGAUGGUACAGG
UACAAUUUACACAGAACUGGAACCACCUUGUAGGUUUGUUACAGACACACCAAAAGGGCCU
AAAGUGAAAUACUUGUACUUCAUCAAGGCUUAAACAACCUAAAUAGAGGUAUGGUGCUGGG
CAGUUUAGCUGCUACAGUACGUCUUCAGGCUGGAAAUGCUACAGAAGUaCCUGCCAAUUCA
ACUGUGCUUUCCUUCUGUGCUUUUGCAGUAGACCCUGCUAAAGCAUAUaAAGGAUUACCUA
GCAAGUGGAGGACAACCAAUCACCAACUGUGUGAAGAUGUUGUGUACACACACUGGUACAG
GACAGGCAAUUACUGUAACACCAGAAGCUAACAUGGACCAAGAGUCCUYUGGUGGUGCUUC
AUGUUGUCUGUAUUGUAGAUGCCACAUUGACCAUCCAAAUCCUAAAGGAYUCUGUGACUUG
AAAGGUAAGUACGUCCAAAUACCUACCACUUGUGCUAAUGACCCAGUGGGUUUUACACUUA
GAAACACAGUCUGUACCGUCUGCGGAAUGUGGAAAGGUUAUGGCUGUAGUUGUGACCAACU
CCGCGAACCCUUGAUGCAGUCUGCGGAUGCAUCAMCGUUUUUAAACGGGUUUGCGGUGUAA
GUGCAGCCCGUCUUACACCGUGCGGCACAGGCACUAGUACUGAUGUCGUCUACAGGGCUUU
UGAUAUUUACAACGAAAAAGUUGCUGGUUYUGCAAAGUUCCUAAAAACUAA

Translation 1
Nucleotide 3-701 ; 233 aa
LEKMADQAMTQMYKQARSEDKRAKVTSAMQTMLFTMLRKLDNDALNNIINNARDGCVPLNIIPLTTAAKLMVV
VPDYGTYKNTCDGNTFTYASALWEIQQVVDADSKIVQLSEINMDNSPNLAWPLIVTALRANSAVKLQNNELSP
VALRQMSCAAGTTQTACTDDNALAYYNNSKGGRFVLALLSDHQDLKWARFPKSDGTGTIYTELEPPCRFVTDT
PKGPKVKYLYFIKA

Translation 2
FKRVCGVSA-ARLTPCGTGTSTDVVYRAFDIYNEKVAGXAKFLK

Alignment 1 of translation 1 sequence
RNA-Dependent RNA Polymerase: bovine coronavirus
Identities = 181/413 (43%),

```
Query: 3    LEKMADQAMTQMYKQARSEDKRAKVTSAMQTMLFTMLRKXXXXXXXXXXXXXRDGCVPLN 182
            LE+MAD A+T MYK+AR  DK++KV SA+QTMLF+M+RK               GCVPLN
Sbjct: 3985 LERMADLALTNMYKEARINDKKSKVVSALQTMLFSMVRKLDNQALNSILDNAVKGCVPLN 4044

Query: 183  IIPLTTAAKLMVVVPDYGTYKNTCDGNTFTYASALWEIQQVVDADSKIVQLSEINMDNSP 362
             IP   A  L ++VPD   Y    D    TYA  +W+IQ + D+D    QL+EI+ D +
Sbjct: 4045 AIPSLAANTLTIIVPDKSVYDQVVDNVYVTYAGNVWQIQTIQDSDGTNKQLNEISDDCN- 4103

Query: 363  NLAWPLIVTALRAN--SAVKLQNNELSPVALRQMSCAAGTTQTACTDDNALAYYNNSKGG 536
               WPL++ A R N  SA  LQNNEL P  L+     +G  QT  T      YYNNS  G
Sbjct: 4104 ---WPLVIIANRHNEVSATVLQNNELMPAKLKTQVVNSGPDQTCNTPTQ--CYYNNSNNG 4158

Query: 537  RFVLALLSDHQDLKWARFPKSDGTGTIYTELEPPCRFVTDTPKGPKVKYLYFIKA*TT*I 716
            + V A+LSD   LK+ +  K DG   +  EL+PPC+F     KG K+KYLYF+K   T
Sbjct: 4159 KIVYAILSDVDGLKYTKILKDDG-NFVVLELDPPCKFTVQDVKGLKIKYLYFVKGCNTLA 4217

Query: 717  EVWCWAV*LLQYVFRL------EMLQKYLPIQLCFPSVLLQ*TLLKHIKDYLASGGQPIT 878
                W V  +    RL      E       + LC  SV  + T L    D++  GG PI
Sbjct: 4218 R--GWVVGTISSTVRLQAGTATEYASNSSILSLCAFSVDPKKTYL----DFIQQGGTPIA 4271

Query: 879  NCVKMLCTHTGTGQAITVTPEANMDQESXGGASCCLYCRCHIDHPNPKGXCDLKGKYVQI 1058
            NCVKMLC H GTG AITV P+A  +Q+S GGAS C+YCR  ++HP+  G C L+GK+VQ+
Sbjct: 4272 NCVKMLCDHAGTGMAITVKPDATTNQDSYGGASVCIYCRARVEHPDVDGLCKLRGKFVQV 4331

Query: 1059 PTTCANDPVGFTLRNTVCTVCGMWKGYGCSCDQLREPLMQSADASXFLNGFAV 1217
            P     DPV + L + VC VCG W+   CSC    +  +QS D + FLNGF V
Sbjct: 4332 PVG-IKDPVSYVLTHDVCQVCGFWRDGSCSCVS-TDTTVQSKDTN-FLNGFGV 4381
```

Alignment 2 of translation 2 sequence
RNA-directed RNA polymerase (ORF1B) [murine hepatitis virus]

Identities = 24/44 (54%),

```
Query: 1199 FKRVCGVSA-ARLTPCGTGTSTDVVYRAFDIYNEKVAGXAKFLK 1327
            FKRV G S  ARL PC +G  TDV  RAFDI N   AG   + K
Sbjct: 18   FKRVRGTSVNARLVPCASGLDTDVQLRAFDICNANRAGIGLYYK 61
```

Fig.2, Contd.

Note that this sequence is not fully in frame.

UGACAUCUUACGCGUAUAUGCUAACUUAGGUGAGCGUGUACGCCAAUCAUUAUUAAAGACU
GUACAAUUCUGCGAUGCUAUGCGUGAUGCAGGCAUUGUAGGCGUACUGACAUUAGAUAAUC
AGGAUCUUAAUGGGAACUGGUACGAUUUCGGUGAUUUCGUACAAGUAGCACCAGGCUGCGG
AGUUCCUAUUGUGGAUUCAUAUUACUCAUUGCUGAUGCCCAUCCUCACUUUGaCUAGGgCA
UUGGCUGCUGAGUCCcAUAUGGAUGCUGAUCUCGCAAAaCCACUUAUUAaGUGGgAUUUGC
UGAAACAUGAUUUUACGGAAGAGAGACUUUGUCUCUUCGACCGUUAUUUUAAAUAUUGGGA
CCAGACAUACCAUCCCAAUUGUAUUAACUGUUUGGAUGAUAGGUGUAUCCUUCAUUGUGCA
AaCUUUAAUGUGUUAUUUUCUACUGUGUUUcCACCUACAAGUUUUGGACCACUAGUAAGAA
AAAUAUUUGUAGAUGGUGUUCCUUCUGUUGUUUCAACUGGAUACCAUUUUCGUGAGUUAGG
AGUCGUACAUAAUCAGGAUGUAAACUUACAUAGCUCGCGUCUCAGUUUCAAGGAACUUUUA
GUGUAUGCUGCUGAUCCAGCUAUGCAUGCAGCUUCUGGCAAUUAAUUGCUAGAUAAACGCA
CUACAUGCUUUUCAGUAGCUCCACUAACAAACAAUGUUGCUUUUCAAACUGUCAAACCCGG
UAAUUUUAAUAAAGACUUUUAUGACUUUGCUGUGUCUAAA

Translation 1

Nucleotide 2 to 652: Frame 2; 217 aa.

DILRVYANLGERVRQSLLKTVQFCDAMRDAGIVGVLTLDNQDLNGNWYDFGDFVQVAPGCGVPIVDSYYSLLM
PILTLTRALAAESHMDADLAKPLIKWDLLKHDFTEERLCLFDRYFKYWDQTYHPNCINCLDDRCILHCANFNV
LFSTVFPPTSFGPLVRKIFVDGVPSVVSTGYHFRELGVVHNQDVNLHSSRLSFKELLVYAADPAMHAASGN

Translation 2

656 to 772: Frame 2; 39 aa

LLDKRTTCFSVAPLTNNVAFQTVKPGNFNKDFYDFAVSK

Alignment

ORFlab polyprotein Murine hepatitis virus

Identities = 157/257 (61%),

```
Query: 2    DILRVYANLGERVRQSLLKTVQFCDAMRDAGIVGVLTLDNQDLNGNWYDFGDFVQVAPGC 181
            DI+ VY  LG    ++LL T +F DA+ +AG+VGVLTLDNQDL G WYDFGDFV+  PGC
Sbjct: 4626 DIINVYKKLGPIFNRALLNTAKFADALVEAGLVGVLTLDNQDLYGQWYDFGDFVKTVPGC 4685

Query: 182  GVPIVDSYYSLLMPILTLTRALAAESHMDADLAKPLIKWDLLKHDFTEERLCLFDRYFKY 361
            GV + DSYYS +MP+LT+  AL +E  ++     +    +DL+++DFT+ +L LF +YFK+
Sbjct: 4686 GVAVADSYYSYMMPMLTMCHALDSELFVNGTYRE----FDLVQYDFTDFKLELFTKYFKH 4741

Query: 362  WDQTYHPNCINCLDDRCILHCANFNVLFSTVFPPTSFGPLVRKIFVDGVPSVVSTGYHFR 541
            W  TYHPN   C DDRCI+HCANFN+LFS V P T FGPLVR+IFVDGVP VVS GYH++
Sbjct: 4742 WSMTYHPNTCECEDDRCIIHCANFNILFSMVLPKTCFGPLVRQIFVDGVPFVVSIGYHYK 4801

Query: 542  ELGVVHNQDVNLHSSRLSFKELLVYAADPAMHAASGN*LLDKRTTCFSVAPLTNNVAFQT 721
            ELGVV N DV+ H  RLS K+LL+YAADPA+H AS + LLD RT CFSVA +T+ V FQT
Sbjct: 4802 ELGVVMNMDVDTHRYRLSLKDLLLYAADPALHVASASALLDLRTCCFSVAAITSGVKFQT 4861

Query: 722  VKPGNFNKDFYDFAVSK 772
            VKPGNFN+DFY+F +SK
Sbjct: 4862 VKPGNFNQDFYEFILSK 4878
```

EMC-7 .

ACCUUCAGAAUUAUGGUGAAAAUGCUGUUAUACCAMAAGGAAUAAUGAUGAAUGUCGCAAAGUAUACUCAACU
GUGUCAAUACUUAAAUACACUUACUUUAGCUGUACCCUACAACAUGAGAGUUAUUCACUUUGGUGCUGGCUCU
GAUAAAGGAGUUGCACCAGGUACAGCUGUGCUCAGACAAUGGUUGCCAACUGGCACACUACUUGUCGAUUCAG
AUCUUAAUGACUUCGUCUCCGACGCAGAUUCUACUUUAAUUGGAGACUGUGCAACAGUACAUACGGCUAAUAA
AUGGGACCUUAUUAUUAGCGAUAUGUAUGACCCUAGGACCAAACAUGUGACAAAAGAGAAUGACUCUAAAGAA
GGGUUUUUCACUUAUCUGUGUGGAUUUAUAAAGCAAAAACUAGCCCUGGGUGGUUCUAUAGCUGUAAAGAUAA
CAGAGCAUUCUUGGAAUGCUGACCUUUACAAGCUUAUGGGCCAUUUCUCAUGGUGGACAGCUUUUGUUACAAA
UGUAAAUGCAUCAUCAUCGGAAGCAUUUUUAAUUGGGGCUAACUAUCUUGGCAAGCCGAAGGAACAAAUUGAU
GGCUAUACCAUGCAUGCUAACUACAUUUUCUGGAGGAACACAAAUCCUAUCCAGUUGUCUUCCUAUUCACUCU
UUGACAUGAGCAAAUUUCCUCUUAAAUUAAGAGGAACUGCUGUAAUGUCUCUUAAGGAGAAUCAAAUCAAUGA
UAUGAUUUAUUCUCUUCUGGAAAAAGGUAGGCUUAUCAUUAGAGAAAACAACAGAGUUGUGGUUUCAAGUGAU
AUUCUUGUUAACAACUAAACGAACAUGUUUAUUUUCUUAUUAUUUCUUACUCUCACUAGUGGUAGUGACCUUG
ACCGGUGCACCACUUUUGAUGAUGUUCAAGCUCCUAAUUACACUCAACAUACUUCAUCUAUGAGGGGGGUUUA
CUAUCCUGAUGAAAUUUUUAGAUCAGACACUCUUUAUUUAACUCAGGAUUUAUUUCUUCCAUUUUAUUCUAAU
GUUACAGGGUUUCAUACUAUUAAUCAUACGUUUGGCAACCCUGUCAUACCUUUUAAGGAUGGUAUUUAUUUUG
CUGCCACAGAGAAAUCAAAUGUUGUCCGUGGUUGGGUUUUUGGUUCUACCAUGAACAACAAGUCACAGUCGGU

Fig.2, Contd.

```
GAUUAUUAUUAACAAUUCUACUAAUGUUGUUAUACGAGCAUGUAACUUUGAAUUGUGUGACAACCCUUUCUUU
GCUGUUUCUAAACCCAUGGGUACACAGACACAUACUAUGAUAUUCGAUAAUGCAUUUAAUUGCACUUUCGAGU
ACAUAUCUGAUGCCUUUUCGCUUGAUGUUUCAGAAAAGUCAGGUAAUUUUAAACACUUACGAGAGUUUGUGUU
UAAAAAUAAAGAUGGGUUUCUCUAUGUUUAUAAGGGCUAUCAACCUAUAGAUGUAGUUCGUGAUCUACCUUCU
GGUUUUAACACUUUGAAACCUAUUUUUAAGUUGCCUCUUGGUAUUAACAUUACAAAUUUUAGAGCCAUUCUUA
CAGCCUUUUCACCUGCUCAAGACAUUUGGGGCACGUCAGCUGCAGCCUAUUUUGUUGGCUAUUUAAAGCCAAC
UACAUUUAUGCUCAAGUAUGAUGAAAAUGGUACAAUCACAGAUGCUGUUGAUUGUUCUCAAAAUCCACUUGCU
GAACUCAAAUGCUCUGUUAAGAGCUUUGAGAUUGACAAAGGAAUUUACCAGACCUCUAAUUUCAGGGUUGUUC
CCUCAGGAGAUGUUGUGAGAUUCCCUAAUAUUACAAACUUGUGUCCUUUUGGAGAGGUUUUUAAUGCUACUAA
AUUCCCUUCUGUCUAUGCAUGGGAGAGAAAAAAAAUUUCUAAUUGUGUUGCUGAUUACUCUGUGCUCUACAAC
UCAACAUUUUUUUCAACCUUUAAGUGCUAUGGCGUUUCUGCCACUAAGUUGAAUGAUCUUUGCUUCUCCAAUG
UCUAUGCAGAUUCUUUUGUAGUCAAGGGAGAUGAUGUAAGACAAAUAGCGCCAGGACAAACUGGUGUUAUUGC
UGAUUAUAAUUAUAAAUUGCCAGAUGAUUUCAUGGGUUGUGUCCUUGCUUGGAAUACUAGGAACAUUGAUGCU
ACUUCAACUGGUAAUUAUAAUUAUAAAUAUAGGUAUCUUAGACAUGGCAAGCUUAGGCCCUUUGAGAGAGACA
UAUCUAAUGUGCCUUUCUCCCCUGAUGGCAAACCUUGCACCCCACCUGCUCUUAAUUGUUAUUGGCCAUUAAA
UGAUUAUGGUUUUUACACCACUACUGGCAUUGGCUACCAACCUUACAGAGUUGUAGUACUUUCUUUUGAACUU
UUAAAUGCACCGGCCACGGUUUGUGGACCAAAAUUAUCCACUGACCUUAUUAAGAACCAGUGUGUCAAUUUUA
AUUUUAAUGGACUCACUGGUACUGGUGUGUUAACUCCUUCUUCAAAGAGAUUUCAACCAUUUCAACAAUUUGG
CCGUGAUGUUUCUGAUUUCACUGAUUCCGUUCGAGAUCCUAAAACAUCUGAAAUAUUAGACAUUUCACCUUGC
UCUUUUGGGGGUGUAAGUGUAAUUACACCUGGAACAAAUGCUUCAUCUGAAGUUGCUGUUCUAUAUCAAGAUG
UUAACUGCACUGAUGUUUCUACAGCAAUUCAUGCAGAUCAACUCACACCAGCUUGGCGCAUAUAUUCUACUGG
AAACAAUGUAUUCCAGACUCAAGCAGGCUGUCUUAUAGGAGCUGAGCAUGUCGACACUUCUUAUGAGUGCGAC
AUUCCUAUUGGAGCUGGCAUUUGUGCUAGUUACCAUACAGUUUCUUUAUUACGUAGUACUAGCCAAAAAUCUA
UUGUGGCUUAUACUAUGUCUUUAGGUGCUGAUAGUUCAAUUGCUUACUCUAAUAACACCAUUGCUAUACCUAC
UAACUUUUCAAUUAGCAUUACUACAGAAGUAAUGCCUGUUUCUAUGGCUAAAACCUCCGUAGAUUGUAAUAUG
UACAUCUGCGGAGAUUCUACUGAAUGUGCUAAUUUGCUUCUCCAAUAUGGUAGCUUUUGCACACAACUAAAUC
GUGCACUCUCGUGGUAUUGCUGCUGAACAGGAUCGCAACACAC
```

Translation 1

Nucleotides 3 to 818: Frame 3   272 aa (orf lab)

```
LQNYGENAVIPQGIMMNVAKYTQLCQYLNTLTLAVPYNMRVIHFGAGSDKGVAPGTAVLRQWLPTGTLLVDSDLNDFVSDA
DSTLIGDCATVHTANKWDLIISDMYDPRTKHVTKENDSKEGFFTYLCGFIKQKLALGGSIAVKITEHSWNADLYKLMGHFS
WWTAFVTNVNASSSEAFLIGANYLGKPKEQIDGYTMHANYIFWRNTNPIQLSSYSLFDMSKFPLKLRGTAVMSLKENQIND
MIYSLLEKGRLIIRENNRVVVSSDILVNN
```

Translation 2

Nucleotide 828 to 3089: Frame 3 756 aa (S protein)

```
MFIFLLFLTLTSGSDLDRCTTFDDVQAPNYTQHTSSMRGVYYPDEIFRSDTLYLTQDLFLPFYSNVTGFHTINHTFGNPVI
PFKDGIYFAATEKSNVVRGWVFGSTMNNKSQSVIIINNSTNVVIRACNFELCDNPFFAVSKPMGTQTHTMIFDNAFNCTFE
YISDAFSLDVSEKSGNFKHLREFVFKNKDGFLYVYKGYQPIDVVRDLPSGFNTLKPIFKLPLGINITNFRAILTAFSPAQD
IWGTSAAAYFVGYLKPTTFMLKYDENGTITDAVDCSQNPLAELKCSVKSFEIDKGIYQTSNFRVVPSGDVVRFPNITNLCP
FGEVFNATKFPSVYAWERKKISNCVADYSVLYNSTFFSTFKCYGVSATKLNDLCFSNVYADSFVVKGDDVRQIAPGQTGVI
ADYNYKLPDDFMGCVLAWNTRNIDATSTGNYNYKYRYLRHGKLRPFERDISNVPFSPDGKPCTPPALNCYWPLNDYGFYTT
TGIGYQPYRVVVLSFELLNAPATVCGPKLSTDLIKNQCVNFNFNGLTGTGVLTPSSKRFQPFQQFGRDVSDFTDSVRDPKT
SEILDISPCSFGGVSVITPGTNASSEVAVLYQDVNCTDVSTAIHADQLTPAWRIYSTGNNVFQTQAGCLIGAEHVDTSYEC
DIPIGAGICASYHTVSLLRSTSQKSIVAYTMSLGADSSIAYSNNTIAIPTNFSISITTEVMPVSMAKTSVDCNMYICGDST
ECANLLLQYGSFCTQLNRALSWYCC
```

Alignment 1 of translation 1

replicase [bovine coronavirus]

Identities = 183/271 (67%),

```
Query: 3     LQNYGENAVIPQGIMMNVAKYTQLCQYLNTLTLAVPYNMRVIHFGAGSDKGVAPGTAVLR 182
             L NYG+   +P G MMNVAKYTQLCQYLNT TLAVP NMRV+H GAGS+KGVAPG+AVLR
Sbjct: 6822  LWNYGKPVTLPTGCMMNVAKYTQLCQYLNTTTLAVPVNMRVLHLGAGSEKGVAPGSAVLR 6881

Query: 183   QWLPTGTLLVDSDLNDFVSDADSTLIGDCATVHTANKWDLIISDMYDPRTKHVTKENDSK 362
             QWLP GT+LVD+DL  FVSD+ +T  GDC T+    +WDLIISDMYDP TK++ + N SK
Sbjct: 6882  QWLPAGTILVDNDLYPFVSDSVATYFGDCITLPFDCQWDLIISDMYDPITKNIGEYNVSK 6941

Query: 363   EGFFTYLCGFIKQKLALGGSIAVKITEHSWNADLYKLMGHFSWWTAFVTNVNASSSEAFL 542
             +GFFTY+C  I+ KLALGGS+A+KITE SWNA+LYKLMG+F++WT F TN NASSSE FL
```

Fig.2,Contd.

```
Sbjct: 6942 DGFFTYICHMIRDKLALGGSVAIKITEFSWNAELYKLMGYFAFWTVFCTNANASSSEGFL 7001

Query: 543  IGANYLGKPKEQIDGYTMHANYIFWRNTNPIQLSSYSLFDMSKFPLKLRGTAVMSLKENQ 722
            IG NYLGKPK +IDG  MHANY+FWRN+        +YSLFDM+KFPLKL GTAV++L+ +Q
Sbjct: 7002 IGINYLGKPKVEIDGNVMHANYLFWRNSTVWNGGAYSLFDMAKFPLKLAGTAVINLRADQ 7061

Query: 723  INDMIYSLLEKGRLIIRENNRVVVSSDILVN 815
            INDM+YSLLEKG+L++R+ N+ V    D LVN
Sbjct: 7062 INDMVYSLLEKGKLLVRDTNKEVFVGDSLVN 7092
```

Alignment 2 (Spike protein of coronavirus)
E2 glycoprotein precursor - murine hepatitis virus (strain JHM); contains spike glycoprotein

Identities = 199/798 (24%), Positives = 314/798 (39%), Gaps = 48/798 (6%)
Frame = +3

```
Query: 828  MFIFLLFLTLTSGSDLDRCTTFDDVQAPNYTQHTSSM-------------RGVYYP-DEI 965
            +F+F+L L    G  D    F  +Q  NY  + +S                RG YY  D +
Sbjct: 2    LFVFILLLPSCLGYIGD----FRCIQTVNYNGNNASAPSISTEAVDVSKGRGTYYVLDRV 57

Query: 966  FRSDTLYLTQDLFLPF----YSNV--TGFHTINHTFGNP--VIPFKDGIYFAATE-KSNV 1118
            + + TL LT    + P    Y N+  TG +T++ T+   P  +   F DGI+       K+N
Sbjct: 58   YLNATLLLTG--YYPVDGSNYRNLALTGTNTLSLTWFKPPFLSEFNDGIFAKVQNLKTNT 115

Query: 1119 VRGW-------VFGSTMNNKXXXXXXXXXXXXXXXXRACNFELCDNPFFAVSKPMGTQTHT 1277
              G        V GS   N                    C + +C  P+    KP
Sbjct: 116  PTGATSYFFPTIVIGSLFGNTSYTVVLEPYNNIIMASVCTYTICQLPY-TPCKP------- 167

Query: 1278 MIFDNAFNCTFEYISDAFSLDVSEKSGNFKHLREFVFKNKDGFLYVY---KGYQPIDVVR 1448
                   N    +   DV        K  R F F       +LY +   +G
Sbjct: 168  -------NTNGNRVIGFWHTDVKPPICLLK--RNFTFNVNAPWLYFHFYQQGGTFYAYYA 218

Query: 1449 DLPSGFNTLKPIFKLPLGINITNFRAILTAFSPAQDIWGTSAAAYFVGYLKPTTFMLKYD 1628
            D PS     L   F + +G  +T +  +    +P     T A  Y+V  L     ++  ++
Sbjct: 219  DKPSATTFL---FSVYIGDILTQYFVLPFICTPTAG--STLAPLYWVTPLLKRQYLFNFN 273

Query: 1629 ENGTITDAVDCSQNPLAELKCSVKSFEIDKGIYQTSNFRVVPSGDVVR-FPNITNLCPFG 1805
            E G IT AVDC+ +  ++E+KC  +S     G+Y  S + V P G V R  PN+ + C
Sbjct: 274  EKGVITSAVDCASSYISEIKCKTQSLLPSTGVYDLSGYTVQPVGVVYRRVPNLPD-CKIE 332

Query: 1806 EVFNATKFPSVYAWERKKISNCVADYSVLYNSTFFSTFKCYGVSATKLNDLCFSNVYADS 1985
            E   A   PS   WER+   NC  + S L          +  C  + A+K+  +CF +V  D
Sbjct: 333  EWLTAKSVPSPLNWERRTFQNCNFNLSSLLRYVQAESLSCNNIDASKVYGMCFGSVSVDK 392

Query: 1986 FVVKGDDVRQIAPGQTGVIADYNYKLPDDFMGCVLAWNTRNIDATSTGNYNYKYRYLRHG 2165
            F +        +  G +G +   NYK+       C L ++     + T  NYN        R+G
Sbjct: 393  FAIPRSRQIDLQIGNSGFLQTANYKIDTAATSCQLYYSLPKNNVT-INNYNPSSWNRRYG 451

Query: 2166 KLRPFERDISNVPFSPDGKPCTPPALNCYWPLNDYGFYTTTGIGYQPYRVVVLSFELLNA 2345
                                           + +ND              R  + +  LLN
Sbjct: 452  -----------------------------FKVND--------------RCQIFANILLNG 468

Query: 2346 --PATVCGPKL---STDLIKNQCVNFNFNGLTGTGVLTP-SSKRFQPFQQFGRDVSDFTD 2507
                  T C   L   +T++   CV ++  G+TG GV    +    +Q    DV+    +
Sbjct: 469  INSGTTCSTDLQLPNTEVATGVCVRYDLYGITGQGVFKEVKADYYNSWQALLYDVNGNLN 528

Query: 2508 SVRDPKTSEILDISPCSFGGVSVITPGTNASSEVAVLYQDVNCTDVSTAIHADQLTPAWR 2687
              RD  T++    I  C   G VS          + E A+LY+++NC+ V T    +  +  P
Sbjct: 529  GFRDLTTNKTYTIRSCYSGRVSAAY--HKEAPEPALLYRNINCSYVFTNNISREENPL-- 584

Query: 2688 IYSTGNNVFQTQAGCLIGAEH--VDTSYECDIPIGAGICASYHTVSLLR---STSQK--S 2846
                  N F +  GC++ A++    +    C++ +GAG+C  Y       R    ST  +  +
Sbjct: 585  ------NYFDSYLGCVVNADNRTDEALPNCNLRMGAGLCVDYSKSRRARRSVSTGYRLTT 638

Query: 2847 IVAYTMSLGADSSIAYSN-NTIAIPTNFSISITTEVMPVSMAKTSVDCNMYICGDSTECA 3023
               Y   L  DS   +    + IPTNF+I     E + +    K ++DC   ++CGD+  C
Sbjct: 639  FEPYMPMLVNDSVQSVGGLYEMQIPTNFTIGHHEEFIQIRAPKVTIDCAAFVCGDNAACR 698

Query: 3024 NLLLQYGSFCTQLNRALS 3077
```

Fig.2, Contd.

```
             L++YGSFC  +N  L+
Sbjct: 699  QQLVEYGSFCDNVNAILN 716
```

RDG1 seq

UUCAAAGCcUUCAAACNUAUGUAACACAACAACUAAUCAGGGMUGcUGAAAUCHCGSCUUCUGCUAAUCUUGC
UGCUACUAAAAUGUCUGAGUGUGUUCUUGGACAAUCAAAAAGAGUUGACUUUUGUGGAAAGGGCUACCACCUU
AUGUCCUUCCCACAAGCAGCCCCGCAUGGUGUUGUCUUCCUACAUGUCACGUAUGUGCCAUCCCAGGAGAGGA
ACUUCACCACAGCGCCAGCAAUUUGUCAUGAAGGCAAAGCAUACUUCCCUCGUGAAGGUGUUUUUGUGUUUAA
UGGCACUUCUUGGUUUAUUACACAGAGGAACUUCUUUUCUCCACAAAUAAUUACUACAGACAAUACAUUUGUC
UCAGGAAAUUGUGAUGUCGUUAUUGGCAUCAUUAACAACACAGUUUAUGAUCCUCUGCAACCUGAGCUUGACU
CAUUCAAAGAAGAGCUGGACAAGUACUUCAAAAAUCAUACAUCACCAGAUGUUGAUCUUGGCGACAUUUCAGG
CAUUAACGCUUCUGUCGUCAACAUUCAAAAAGAAAUUGACCGCCUCAAUGAGGUCGCUAAAAAUUUAAAUGAA
UCACUCAUUGACCUUCAAGAAUUGGGAAAAUAUGAGCAAUAUAUUAAGUGgCCCUGGUACGUCUGGGU

Translation 1
Nucleotides 3 to 650: Frame 3; 216 aa
QSLQXYVTQQLIRXAEIXXSANLAATKMSECVLGQSKRVDFCGKGYHLMSFPQAAPHGVVFLHVTYVPSQERNFTTAPAIC
HEGKAYFPREGVFVFNGTSWFITQRNFFSPQIITTDNTFVSGNCDVVIGIINNTVYDPLQPELDSFKEELDKYFKNHTSPD
VDLGDISGINASVVNIQKEIDRLNEVAKNLNESLIDLQELGKYEQYIKWPWYVW

Translation 2
Nucleotides 37 to 339: Frame 1; 101 aa
SGXLKXXLLLILLLLKCLSVFLDNQKELTFVERATTLCPSHKQPRMVLSSYMSRMCHPRRGTSPQRQQFVMKAKHTSLVKV
FLCLMALLGLLHRGTSFLHK

Translation 3
Nucleotides 343 to 576: Frame 1; 78 aa
LLQTIHLSQEIVMSLLASLTTQFMILCNLSLTHSKKSWTSTSKIIHHQMLILATFQALTLLSSTFKKKLTASMRSLKI

Alignment of translation 1

S glycoprotein [murine hepatitis virus]
Length = 1376

Identities = 86/218 (39%), Positives = 129/218 (59%), Gaps = 3/218 (1%)
Frame = +3

```
Query: 6    SLQTYVTQQLIRXAEIXXSANLAATKMSECVLGQSKRVDFCGKGYHLMSFPQAAPHGVVF 185
            +L  Y+++QL      I  SA  A  K++ECV  Q+ R++FCG G H++S  Q AP+G+ F
Sbjct: 1105 ALNAYISKQLSDSTLIKFSAAQAIEKVNECVKSQTTRINFCGNGNHILSLVQNAPYGLYF 1164

Query: 186  LHVTYVPSQERNFTTAPAICHEG-KAYFPREGVFVFNGTSWFITQRNFFSPQIITTDNTF 362
            +H +YVP+          +P +C  G +   P+ G FV +    W  T  +++ P+ IT  N+
Sbjct: 1165 IHFSYVPTSFTTANVSPGLCISGDRGLAPKAGYFVQDDGEWKFTGSSYYYPEPITDKNSV 1224

Query: 363  VSGNCDVVIGIINNTVYDPLQPELDSFKEELDKYFKNHTS--PDVDLGDISGINASVVNI 536
            V  +C V         + +   P L  FKEELDK+FKN TS  PD+ L D    +N + +++
Sbjct: 1225 VMSSCSVNYTKAPEVLLNSSIPNLPDFKEELDKWFKNQTSIAPDLSL-DFEKLNVTFLDL 1283

Query: 537  QKEIDRLNEVAKNLNESLIDLQELGKYEQYIKWPWYVW 650
              E++R+ E  K LNES I+L+E+G YE Y+KWPWYVW
Sbjct: 1284 SDEMNRIQEAIKKLNESYINLKEVGTYEMYVKWPWYVW 1321
```

EMC-8

AGGCCAAAACAGCGCCGACCCCAAGGUUUACCCAAUAAUACUGCGUCUUGGUUCACAGCUCUCACUCAGCAUG
GCAAGGAGGAACUUAGAUUCCCUCGAGGCCAGGGCGUUCCAAUCAACACCAAUAGUGGUCCAGAUGACCAAAU
UGGCUACUACCGAAGAGCUACCCGACGAGUUCGUGGUGGUGACGGCAAAAUGAAAGAGCUCAGCCCCAGAUGG
UACUUCUAUUACCUAGGAACUGGCCCAGAAGCUUCACUUCCCUACGGCGCUAACAAAGAAGGCAUCGUAUGGG
UUGCAACUGAGGGAGCCUUGAAUACACCCAAAGACCACAUUGGCACCCGCAAUCCUAAUAACAAUGUUGCC

Translation
Nucleotides 1 to 363: Frame 1; 121 aa
RPKQRRPQGLPNNTASWFTALTQHGKEELRFPRGQGVPINTNSGPDDQIGYYRRATRRVRGGDGKMKELSPRWYFYYLGTG
PEASLPYGANKEGIVWVATEGALNTPKDHIGTRNPNNNXA

Fig.2, Contd.

Alignment

nucleocapsid protein - bovine coronavirus (strain Mebus)

```
 Identities = 55/129 (42%),

Query: 1   RPKQRRPQGLPNNTA-----SWFTALTQHGK-EELRFPRGQGVPINTNSGPDDQIGYYRR 162
           +PKQ    LP+         SWF+ +TQ  K +E  F  GQGVPI         +  GY+ R
Sbjct: 44  QPKQTATSQLPSGGNVVPYYSWFSGITQFQKGKEFEFAEGQGVPIAPGVPATEAKGYWYR 103

Query: 163 ATRR-VRGGDGKMKELSPRWYFYYLGTGPEASLPYGANKEGIVWVATEGA-LNTPKDHIG 336
             RR  +  DG  ++L PRWYFYYLGTGP A   YG + +G+ WVA+  A +NTP D I
Sbjct: 104 HNRRSFKTADGNQRQLLPRWYFYYLGTGPHAKDQYGTDIDGVFWVASNQADVNTPAD-IL 162

Query: 337 TRNPNNNXA 363
             R+P+++ A
Sbjct: 163 DRDPSSDEA 171
```

<u>EMC-11</u>: unknown sequence

UUGCAUACCGCAAUGUUCUUCUUCGUAAGAACGGUaAUAAGGGAGCCGGUGGUCAUAGCUgUGGCAUGAUCUA
AAGUCUUAUGACUUAGGUGACGAGCUUGGCACUGAUCCCAUUGAAGAUUAUGAACAAAACUGGAACACUAAGC
AUGGCAGUGGUGCACUCCGUGAACUCACUCGUGAGCUCAAUGGAGGUGCAGUCACUCGCUAUGUCGACAACAA
UUUCUGUGGCCCAGAUGGGUACCCUCUUGAUUGCAUCAAAGAUUUUCUCGCACGCGCGGGCAAGUCAAUGUGC
ACUCUUUCCGAACAACUUGAUUACAUCGAGUCGaAGAGAGGUGUCUACUGCUGCCGUGACCAUGAGCAUGAAA
UUGCCUgGGUUCACUGAGCGCUCUGAUAAGAGCUACGAGCACCAGACACCCUUCGaAAUUAAGAGUGCCAAGA
AAaUUGACACUUUCAAAAGGGGAAUGCCCCAAAGCUUGUGUUUCCUCUUAACUCAAAAGUCAAAGUCAUUCAA
CCACGUGUUGAAAAGAAAAAGACUGAGGGUUUCAUGGGGCGUAUACGCUCUGUGUACCCUGUUGCAUCUCCAC
AGGAGUGUAACAAUAUGCACUUGUCUACCUUGAUGAAAUGUAAUCAUUGCGAUGAAGCUUCAUGGCAGACGUG
CGACUUUCUGAAAGCCACUUGUGAACAUUGUGGCACUGAAAAUUUAGUUAUUGAAGGACCUAGUACAUGUGGG
UACCUACCUACUAAUGCUGUAGUGAAAAUGCCAUGUCCUGCCUGUCAAGACCCAGAGAUUGGACCUGAGCAUA
GUGUUGCAGAUUAUCACAACCACUCAAACAUUGAAACUCGACUCCGCAAGGGAGGUAGGACUAGAUGUUUUGG
AGGCUGUGUGUUUGCCUAUGUUGGCUGCUAUAAUAAGCGUGCCUACUGGGUUCCUCGUGCUAGUGCUGAUAUU
GGCUCAGGCCAUACUGGCAUUACUGGUGACAAUGUGGAGACCUUGAAUGAGGAUCUCCUUGAGAUACUGAGUC
GUGAACGUGUUAACAUUAACAUUGUUGGCGAUUUUCAUUUGAAUGAAGAGGUUGCCAUCAYUUUGGCAUCYUU
CUCUGCUUCUACAAGUGCCUUUAUUGACACUAUAAAGAGUCUUGAUUACAAGUCUUUCAAAACCAUUGUUGAG
UCCUGCGGUAACUAUAAAGUUACCAAGGGAAAGCCCGUAAAAGGUGCUUGGAACAUUGGACAACAGAGAUCAG
UUUUAACACCACUGUGUGGUUUUCCCUCACAGGCUGCUGGUGUUAUCAGAUCAAUUUUUGCGCGCACACUUGA
UGCAGCAAACCACUCAAUUCCUGAUUUGCAAAGAGCAGCUGUCACCAUACUUGAUGGUAUUUCUGAACAGUCA
UUACGUCUUGUCGACGCCAUGGUUUAUACUUCAGACCUGCUCACCAACAGUGUCAUUAUUAUGGCAUAUGUAA
CUGGUGGUCUUGUACAACAGACU

Translation of putative open reading frames

```
>~out: 78 to 1: Frame -2        26 aa
DFRSCHSYDHRLPYYRSYEEEHCGMQ
>~out: 59 to 379: Frame 2        107 aa
LWHDLKSYDLGDELGTDPIEDYEQNWNTKHGSGALRELTRELNGGAVTRYVDNNFCGPDGYPLDCIKDFLARAGKSMCTLS
EQLDYIESKRGVYCCRDHEHEIAWVH
>~out: 283 to 89: Frame -1        65 aa
LARACEKIFDAIKRVPIWATEIVVDIASDCTSIELTSEFTECTTAMLSVPVLFIIFNGISAKLVT
>~out: 90 to 614: Frame 3        175 aa
VTSLALIPLKIMNKTGTLSMAVVHSVNSLVSSMEVQSLAMSTTISVAQMGTLLIASKIFSHARASQCALFPNNLITSSRRE
VSTAAVTMSMKLPGFTERSDKSYEHQTPFEIKSAKKIDTFKRGMPQSLCFLLTQKSKSFNHVLKRKRLRVSWGVYALCTLL
HLHRSVTICTCLP
>~out: 204 to 124: Frame -2        27 aa
RVTAPPLSSRVSSRSAPLPCLVFQFCS
>~out: 312 to 208: Frame -2        35 aa
SSCSERVHIDLPARARKSLMQSRGYPSGPQKLLST
>~out: 485 to 258: Frame -3        76 aa
EETQALGHSPFESVNFLGTLNFEGCLVLVALIRALSEPRQFHAHGHGSSRHLSSTRCNQVVRKECTLTCPRVRENL
>~out: 397 to 287: Frame -1        37 aa
LLSERSVNPGNFMLMVTAAVDTSLRLDVIKLFGKSAH
>~out: 364 to 486: Frame 1        41 aa
NCLGSLSALIRATSTRHPSKLRVPRKLTLSKGECPKACVSS
>~out: 490 to 401: Frame -1        30 aa
VKRKHKLWGIPLLKVSIFLALLISKGVWCS
>~out: 446 to 1483: Frame 2        346 aa
```

Fig.2, Contd.

HFQKGNAPKLVFPLNSKVKVIQPRVEKKKTEGFMGRIRSVYPVASPQECNNMHLSTLMKCNHCDEASWQTCDFLKATCFHC
GTENLVIEGPSTCGYLPTNAVVKMPCPACQDPEIGPEHSVADYHNHSNIETRLRKGGRTRCFGGCVFAYVGCYNKRAYWVP
RASADIGSGHTGITGDNVETLNEDLLEILSRERVNINIVGDFHLNEEVAIXLAXFSASTSAFIDTIKSLDYKSFKTIVESC
GNYKVTKGKPVKGAWNIGQQRSVLTPLCGFPSQAAGVIRSIFARTLDAANHSIPDLQRAAVTILDGISEQSLRLVDAMVYT
SDLLTNSVIIMAYVTGGLVQQT
>~out: 643 to 494: Frame -1 50 aa
SFIAMITFHQGRQVHIVTLLWRCNRVHRAYTPHETLSLFLFNTWLNDFDF
>~out: 627 to 511: Frame -2 39 aa
LHFIKVDKCILLHSCGDATGYTERIRPMKPSVFFFSTRG
>~out: 704 to 612: Frame -3 31 aa
LNFQCHNVHKWLSESRTSAMKLHRNDYISSR
>~out: 774 to 631: Frame -2 48 aa
QAGHGIFTTALVGRYPHVLGPSITKFSVPQCSQVAFRKSHVCHEASSQ
>~out: 826 to 737: Frame -1 30 aa
VVVIICNTMLRSNLWVLTGRTWHFHYSISR
>~out: 863 to 744: Frame -3 40 aa
SYLPCGVEFQCLSGCDNLQHYAQVQSLGLDRQDMAFSLQH
>~out: 756 to 992: Frame 3 79 aa
KCHVLPVKTQRLDLSIVLQIITTTQTLKLDSAREVGLDVLEAVCLPMLAAIISVPTGFLVLVLILAQAILALLVTMWRP
>~out: 952 to 830: Frame -1 41 aa
ANISTSTRNPVGTLIIAANIGKHTASKTSSPTSLAESSFNV
>~out: 1056 to 922: Frame -2 45 aa
KSPTMLMLTRSRLSISRRSSFKVSTLSPVMPVWPEPISALARGTQ
>~out: 1237 to 956: Frame -1 94 aa
SLLSNVPSTFYGLSLGNFIVTAGLNNGFERLVIKTLYSVNKGTCRSREXCQXDGNLFIQMKIANNVNVNTFTTQYLKEILI
QGLHIVTSNASMA
>~out: 1140 to 1060: Frame -2 27 aa
SRLFIVSIKALVEAEXDAKXMATSSFK
>~out: 1131 to 1205: Frame 3 25 aa
RVLITSLSKPLLSPAVTIKLPRESP
>~out: 1410 to 1183: Frame -2 76 aa
TMASTRRNDCSEIPSSMVTAALCKSGIEWFAASSVRAKIDLITPAACEGKPHSGVKTDLCCPMFQAPFTGFPLVTL
>~out: 1186 to 1311: Frame 1 42 aa
SYQGKARKRCLEHWTTEISFNTTVWFSLTGCWCYQINFCAHT
>~out: 1283 to 1191: Frame -3 31 aa
HQQPVRENHTVVLKLISVVQCSKHLLRAFPW
>~out: 1248 to 1457: Frame 3 70 aa
HHCVVFPHRLLVLSDQFLRAHLMQQTTQFLICKEQLSPYLMVFLNSHYVLSTPWFILQTCSPTVSLLWHM
>~out: 1381 to 1482: Frame 1 34 aa
TVITSCRRHGLYFRPAHQQCHYYGICNWWSCTTD

EMC12: unknown sequence
UGCUUGCUCAUGCUGAAGAGACAAGAAAAUUAAUGCCUAUAUGCAUGGAUGUUAGAGCCAU
AAUGGCAACCAUCCAACGUAAGUAUAAAGGAAUUAAAAUUCAAGAGGGCAUCGUUGACUAU
GGUGUCCGAUUCUUCUUUUAUACUAGUAAAGAGCCUGUAGCUUCUAUUAUUACGAAGCUGA
ACUCUCUAAAUGAGCCGCUUGUCACAAUGCCAAUUGGUUAUGUGACACAUGGUUUUAAUCU
UGAAGAGGCUGCGCGCUGUAUGCGUUCUCUUAAAGCUCCUGCCGUAGUGUCAGUAUCAUCA
CCAGAUGCUGUUACUACAUAUAAUGGAUACCUCACUUCGUCAUCAAAGACAUCUGAGGAGC
ACUUUGUAGAAACAGUUUCUUUGGCUGGCUCUUACAGAGAUUGGUCCUAUUCAGGACAGCG
UACAGAGUUAGGUGUUGAA

Translation of putative open reading frames
>~out: 3 to 446: Frame 3 148 aa
LAHAEETRKLMPICMDVRAIMATIQRKYKGIKIQEGIVDYGVRFFFYTSKEPVASIITKLNSLNEPLVTMPIGYVTHGFNL
EEAARCMRSLKAPAVVSVSSPDAVTTYNGYLTSSSKTSEEHFVETVSLAGSYRDWSYSGQRTELGVE
>~out: 100 to 11: Frame -2 30 aa
ILIPLYLRWMVAIMALTSMHIGINFLVSSA
>~out: 188 to 33: Frame -1 52 aa
RVQLRNNRSYRLFTSIKEESDTIVNDALLNFNSFILTLDGCHYGSNIHAYRH
>~out: 64 to 159: Frame 1 32 aa
WQPSNVSIKELKFKRASLTMVSDSSFILVKSL
>~out: 220 to 143: Frame -2 26 aa
PIGIVTSGSFREFSFVIIEATGSLLV
>~out: 293 to 192: Frame -1 34 aa
HYGRSFKRTHTARSLFKIKTMCHITNWHCDKRLI
>~out: 397 to 224: Frame -2 58 aa
EPAKETVSTKCSSDVFDDEVRYPLYVVTASGDDTDTTAGALRERIQRAASSRLKPCVT
>~out: 229 to 288: Frame 1 20 aa

Fig.2, Contd.

HMVLILKRLRAVCVLLKLLP
>~out: 292 to 372: Frame 1 27 aa
CQYHHQMLLLHIMDTSLRHQRHLRSTL
>~out: 444 to 340: Frame -3 35 aa
QHLTLYAVLNRTNLCKSQPKKLFLQSAPQMSLMTK
>~out: 416 to 351: Frame -1 22 aa
IGPISVRASQRNCFYKVLLRCL
>~out: 365 to 445: Frame 2 27 aa
GALCRNSFFGWLLQRLVLFRTAYRVRC
>~out: 376 to 435: Frame 1 20 aa
KQFLWLALTEIGPIQDSVQS

Figure 3.

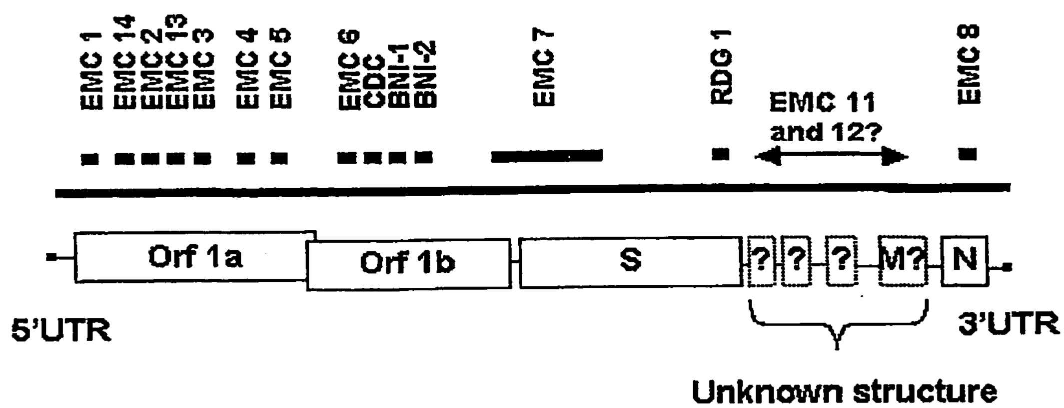
EMC 1
EMC 14
EMC 2
EMC 13
EMC 3
EMC 4
EMC 5
EMC 6
CDC
BNI-1
BNI-2
EMC 7
RDG 1
EMC 11
and 12?
EMC 8
Orf 1a
Orf 1b
S
?
?
?
M?
N
5'UTR
3'UTR
Unknown structure

Figure 4.

Comparison of N-termini of the S proteins of the group 2 coronaviruses

HCV OC43	MFLILLISLPTAFAVIG**DL-KCTT**VSINDID
MHV A59	MLFVFILFLPSCLGYIG**D**F-**RC**IQLVNSNGA
BCV	MFLILLISLPMAFAVIG**DL-KCTT**VSINDVD
SARS	MF-IFLLFL–TLTSG-SDLDRCTTFDDVQAP

Figure 5.

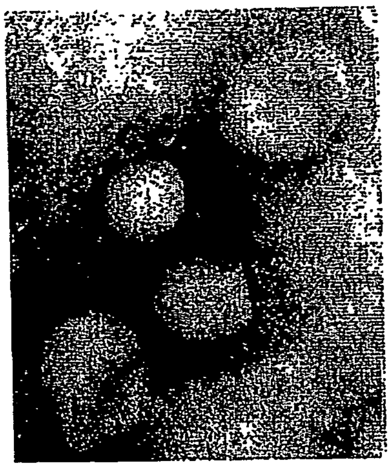

Figure 6.